ECE 2521: Analysis of Stochastic Processes

Lecture 9

Department of Electrical and Computer Engineering
University of Pittsburgh

November, 10^{th} 2021

azime

$X_1, X_2, \dots, X_N \longrightarrow N$ independent measurements of X
(iid)

Example:
$X_i = v + N_i$ (measuring constant voltage with noise added)

<u>sample mean</u> : $M_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

estimator of $E[X]$ $\longrightarrow$ How good is this estimator?

$E[M_n] = \mu$ $\longrightarrow$ should be good estimate of mean

$\longrightarrow$ shouldn't vary too much

1. $E[M_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$

$E[X_i] = E[X] = \mu$ for all i
because iid

unbiased estimator for μ

$$E[(M_n - \mu)^2] = E[(M_n - E[M_n])^2]$$

$$S_n = X_1 + \dots + X_n \longrightarrow M_n = \frac{S_n}{n}$$

$$Var(M_n) = \frac{1}{n^2} Var(S_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Chebyshev Inequality:

$$P[|M_n - E[M_n]| \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow \boxed{P[|M_n - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{n\varepsilon^2}}$$

where ε: error

Continuing example: $X_i = v + N_i$ $N_i \sim (0, 1\ \mu V)$

How many measurements required such that $\varepsilon = 1\ \mu V$ with probabality of 0.99?

$$1 - \frac{\sigma^2}{n\varepsilon^2} = 1 - \frac{1}{n} = 0.99 \Rightarrow n = 100$$

Weak law of large numbers

$$\lim_{n\to\infty} P\left[|M_n - \mu| < \varepsilon\right] = 1 \qquad \varepsilon > 0$$

$X_1, \ldots, X_N \longrightarrow$ iid RVs with finite mean $E[X] = \mu$

meaning: for large enough N, the sample mean will be close to the true mean with high probability

weak because it doesn't make a connection between M_n and n

Strong law of large numbers

$X_1, X_2, X_3, \ldots$

sequence of sample means $M_1, M_2, \ldots, M_J, M_{J+1}$

$x_1 - x_j$

$$P\left[\lim_{n\to\infty} M_n = \mu\right] = 1$$

this convergence will happen eventually and certainly (to the mean)

now n could be t instead, for time

Central Limit Theorem

$x_1, x_2, \dots, x_N$ with finite mean μ, finite variance
(iid)

$\longrightarrow S_n = x_1 + x_2 + \dots + x_n$
$\downarrow$
Gaussian
RV

(constant)
$M_n \longrightarrow \mu$ as $n \longrightarrow \infty$

$X_1, X_2, \dots, X$ (random variable)

$X_n \longrightarrow X$ as $n \longrightarrow \infty$

Sequences of Random Variables

Deterministic sequence you can access by n

- Let $X_1, X_2 \ldots, X_n, \ldots$ be a sequence of random variables such that each X_i maps points in sample space S to $\mathbb{R}$. That is, for every $\xi_j \in S$, $X_1(\xi_j), X_2(\xi_j), \ldots, X_n(\xi_j), \ldots$ is a sequence of real numbers.

Each outcome

	X_1	X_2	X_3	...	X_n	...	X
ξ_i	$X_1(\xi_1)$	$X_2(\xi_1)$	$X_3(\xi_1)$	...	$X_n(\xi_1)$	...	$X(\xi_1)$
ξ_2	$X_1(\xi_2)$	$X_2(\xi_2)$	$X_3(\xi_2)$	...	$X_n(\xi_2)$	...	$X(\xi_2)$
ξ_3	$X_1(\xi_3)$	$X_2(\xi_3)$	$X_3(\xi_3)$	...	$X_n(\xi_3)$	...	$X(\xi_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Note: X_i is a function not a real number.

Relevant for stochastic processes, which change over time

- $X_n \to X$ (X_n convergence to X)

Example

- A fair coin toss is tossed once: $S = \{H, T\}$
- Define a sequence of random variables $X_1, X_2, X_3, \ldots$ as follows:

$$X_n(\xi) = \begin{cases} \frac{1}{n+1} & \text{if } \xi = H \\ 1 & \text{if } \xi = T \end{cases}$$

$P(X_1 = 1, X_2 = 1) = P(T) = \frac{1}{2}$
$P(X_1 = 1)P(X_2 = 1) = \frac{1}{4}$
NO

1. Are the X_i's independent?
2. Find the PMF and CDF of X_n for $n = 1, 2, 3, \ldots$
3. As $n \to \infty$, what does the CDF of X look like?

$$X_1(\xi) = \begin{cases} \frac{1}{2} & \text{if } \xi = H \\ 1 & \text{if } \xi = T \end{cases}$$

$$X_2(\xi) = \begin{cases} \frac{1}{3} & \text{if } \xi = H \\ 1 & \text{if } \xi = T \end{cases}$$

$$\vdots$$

(a) $P(X_1=1,\ X_2=1) = P(T) = \frac{1}{2}$

$P(X_1=1) \cdot P(X_2=1) = P(T) \cdot P(T) = \frac{1}{4}$

X_i's are <u>NOT</u> independent – values are determined by same coin toss

(b) PMF: $p_{X_n}(x) = P(X_n = x) = \begin{cases} \frac{1}{2}, & x = \frac{1}{n+1} \\ \frac{1}{2}, & x = 1 \end{cases}$

CDF: $F_{X_n}(x) = P(X_n \leq x) = \begin{cases} 1 & x \geq 1 \\ \frac{1}{2} & \frac{1}{n+1} \leq x < 1 \\ 0 & x < \frac{1}{n+1} \end{cases}$

(c) As $n \to \infty$

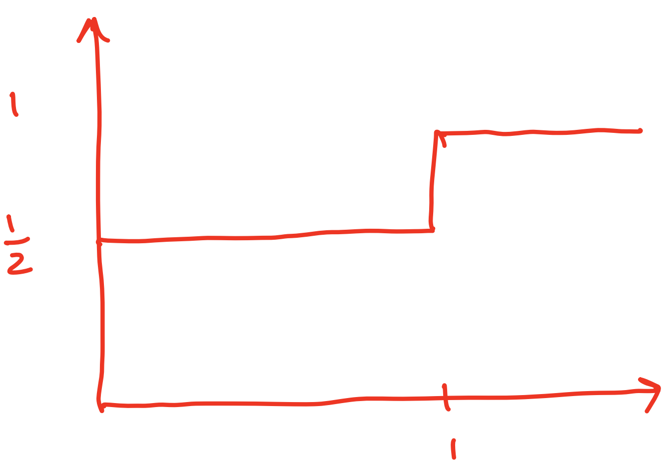
$$F_{X_n}(x) = \begin{cases} 1 & x \geq 1 \\ \frac{1}{2} & 0 \leq x < 1 \\ 0 & x < 0 \end{cases}$$

Does not converge to a fixed value.
CDF approximates CDF of Bernoulli random variable

$$X_\infty(\xi) = \begin{cases} 0 & \text{if } \xi = H \\ 1 & \text{if } \xi = T \end{cases}$$

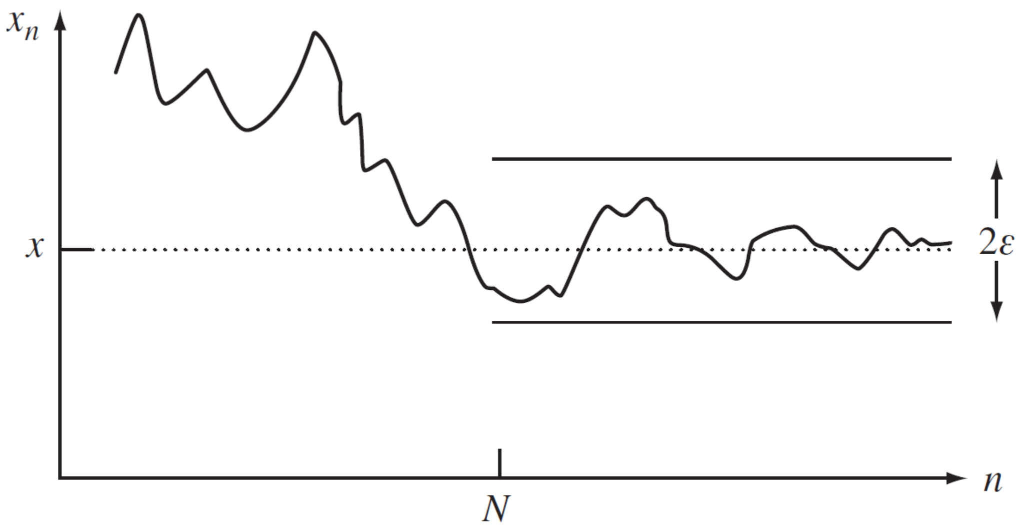
$n \to \infty$ $$F_{X_n}(x) = \begin{cases} 1 & x \geq 1 \\ \frac{1}{2} & 0 \leq x \leq 1 \\ 0 & x < 0 \end{cases}$$

$F_X(x)$ as $n \to \infty$



Convergence of Real Sequences

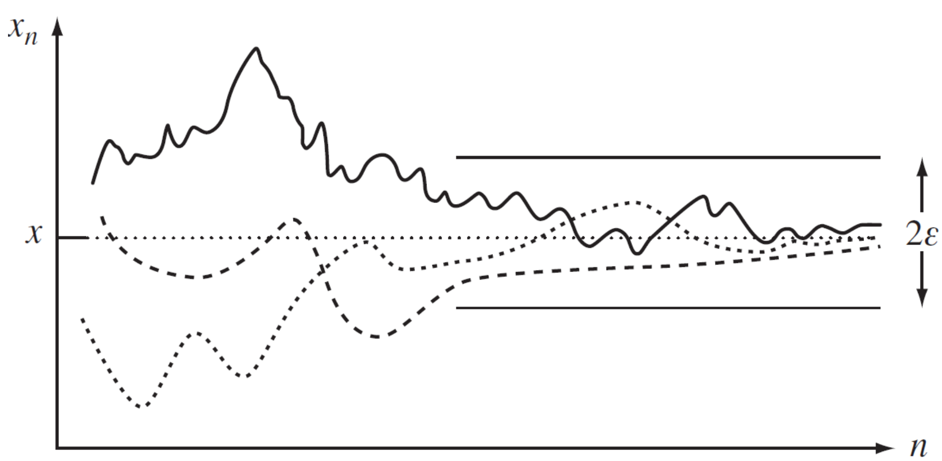
- $\lim_{n\to\infty} x_n = x \iff$ for any $\epsilon > 0$ there exists an $N \in \mathbb{Z}^+$ such that for any $n \geq N$, we have $|x_n - x| < \epsilon$.



Types of Convergence for Random Sequences

1. **Convergence everywhere (surely):** sequence for any outcome will converge to a value
   - $\lim_{n\to\infty} X_n(\xi) = X(\xi)$ for any $\xi \in S$.
2. **Convergence almost everywhere (almost surely):**
   - $P\left[\xi : \lim_{n\to\infty} X_n(\xi) = X(\xi)\right] = 1.$
   - Example: Strong law of large numbers

sequences that don't converge are part of outcomes that do not happen



Example

- Let ξ be selected at random from the interval $S = [0, 1]$, where we assume that the probability that ξ is in a subinterval of S is equal to the length of the subinterval. uniform distribution
- For $n = 1, 2, \ldots$, we define the following five sequences of random variables:

  Can use s instead of ξ

  (1) $U_n = \dfrac{\xi}{n}$
  (2) $V_n = \xi\left(1 - \dfrac{1}{n}\right)$
  (3) $W_n = \xi e^n$
  (4) $Y_n = \cos 2\pi n\xi$
  (5) $Z_n = e^{-n(n\xi-1)}$
- Which of these sequences converge surely? almost surely?
- Identify the limiting random variable.

1. Does converge → goes to 0 as $n \to \infty$

2. $n \to \infty$, $V_n \to 1$ Surely Surely

3. ξe^n, $n \to \infty$, $W \to \infty$, doesn't converge

4. $Y_n = \cos(2\pi n \xi)$: doesn't converge
because it's oscillating

5. $Z_n = e^{-n(n\xi - 1)}$: almost surely, when
$\xi = 0$, $Z_n = e^n$ which diverges, but
for all other $\xi > 0$ it becomes $e^{-n^2\xi + n}$
which goes to 0 as $n \to \infty$

Types of Convergence for Random Sequences

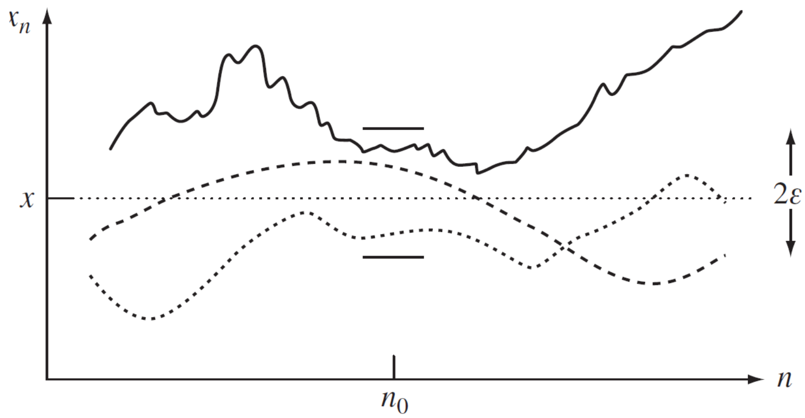
Don't know prob law

3 **Convergence in the mean-square sense:** for $\mathcal{E}$

- $\lim_{n\to\infty} E\left[|X_n - X|^2\right] = 0.$

4 **Convergence in probability:**

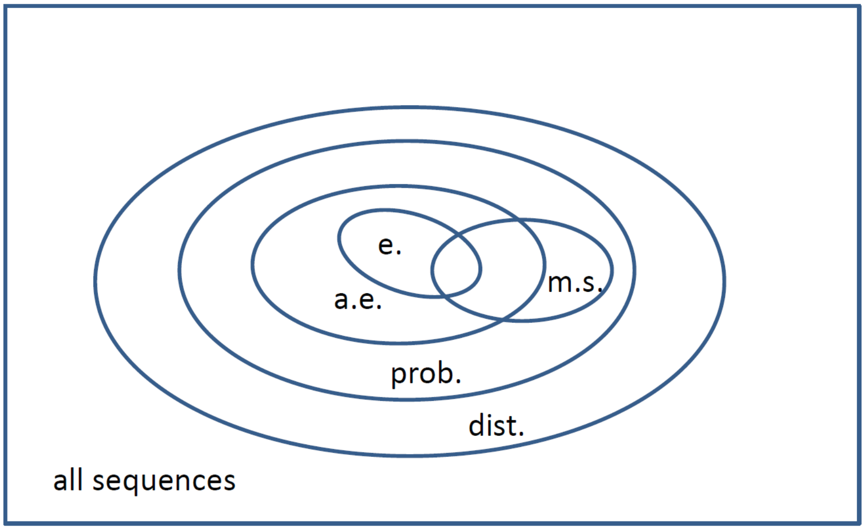
- $\lim_{n\to\infty} P\left[|X_n - X| > \epsilon\right] = 0$ for any $\epsilon > 0.$ ✶
- Example: Weak law of large numbers.



Types of Convergence for Random Sequences

5 **Convergence in distribution:**

- $\lim_{n\to\infty} F_{X_n}(\alpha) = F_X(\alpha)$ at every point α where the cumulative distribution function F is continuous.
- Convergence in distribution $\iff$ $\lim_{n\to\infty} \Phi_{X_n}(s) \to \Phi_X(s)$.
- Example: Central limit theorem



Propositions

conv. in MSS $\Longrightarrow$ conv. in probability

- **Proposition 1:** Convergence in mean-square sense implies convergence in probability, but the converse does not hold.

- **Proposition 2:** Convergence almost everywhere does not imply convergence in the mean-square sense.

Sample Mean Estimation

- Let X be a random variable with mean $E[X] = \mu$ (unknown) and variance σ^2 (finite)
- Let $X_1, X_2, \ldots, X_n$ be a sequence of n independent repeated measurements of X
- The **sample mean** of this sequence is used to estimate μ:

$$M_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

- We will consider $E[M_n]$ and $Var[M_n]$ to assess the effectiveness of M_n as an estimator for μ:
  (1) A good estimator should give the correct value on average (i.e. $E[M_n] = \mu$ to be an **unbiased estimator**)
  (2) The estimator should not vary much about the correct value: the mean square error should be small (i.e. $Var[M_n] \to 0$ as $n \to \infty$)

Weak Law of Large Numbers

- Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of i.i.d random variables with finite mean $E[X_i] = \mu$ satisfying

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = 0 \quad \text{(finite variance)}$$

- Then $M_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to \mu$ in probability:

$$\lim_{n\to\infty} P\left[|M_n - \mu| > \epsilon\right] = 0 \text{ for any } \epsilon > 0$$

- **Interpretation:** For a large enough (and fixed) number of samples n, the sample mean using n samples will be close to the μ with high probability.

Strong Law of Large Nnumbers

- Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of i.i.d random variables with finite mean $E[X_i] = \mu$ and finite variance.
- Then $M_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to \mu$ almost everywhere:

$$P\left[\lim_{n\to\infty} M_n = \mu\right] = 1$$

- **Interpretation:** Every sequence of sample mean calculations will eventually approach and stay close to μ with probability 1.

Central Limit Theorem

- The central limit theorem explains why Gaussian random variables arise in so many practical applications.
- Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d random variables, each with mean μ and variance σ^2. Let $W_n = X_1 + \cdots + X_n$. The mean and variance of W_n is given by $n\mu$ and $n\sigma^2$. We define the *standardized* random variable:

$$Z_n = \frac{W_n - n\mu}{\sqrt{\sigma^2 n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{\sigma^2 n}}$$

- One can show that

$$\begin{aligned} E\left[Z_n\right] &= \frac{E\left[X_1 + \cdots + X_n\right] - n\mu}{\sqrt{\sigma^2 n}} = 0 \\ \mathrm{Var}\left[Z_n\right] &= E\left[Z_n^2\right] = \frac{\mathrm{Var}\left[X_1 + \cdots + X_n\right]}{\sigma^2 n} = 1 \end{aligned}$$

- The random variable Z_n is called **standardized** since for all n: $E\left[Z_n\right] = 0$, and $\mathrm{Var}\left[Z_n\right] = 1$.

Central Limit Theorem

Given $X_1, X_2, \ldots$ a sequence of i.i.d random variables with expected value μ and variance σ^2, the CDF of $Z_n = \frac{\sum_{i=0}^{n} X_i - n\mu}{\sqrt{\sigma^2 n}}$ converges to the standard normal CDF:

$$\lim_{n \to \infty} F_{Z_n}(z) = \Phi(z)$$

where the standard normal CDF: $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$ is available from the normal tables.

- What exactly does the Central Limit Theorem say?
  The CDF of Z_n converges to the standard normal CDF (note that this is not a statement about the convergence of PDFs or PMFs)

Central Limit Theorem Approximation

- Let $W_n = X_1 + \cdots + X_n$ be the sum of i.i.d random variables with mean μ and variance σ^2.
- The Central Limit Theorem **approximation** to the CDF of W_n is:

$$F_{W_n}(w) \approx \Phi(\frac{w - n\mu}{\sqrt{n\sigma^2}})$$

Normalized version

How to use the Central Limit Theorem?

1. Express $W_n = X_1 + \cdots + X_n$ in terms of Z_n

$$W_n = \sqrt{n\sigma^2} Z_n + n\mu$$

2. The CDF of W_n can now be expressed in terms of the CDF of Z_n

$$\begin{aligned} F_{W_n}(w) &= P[W_n \leq w] = P\left[\sqrt{n\sigma^2} Z_n + n\mu \leq w\right] \\ &= F_{Z_n}(\frac{w - n\mu}{\sqrt{n\sigma^2}}) \end{aligned}$$

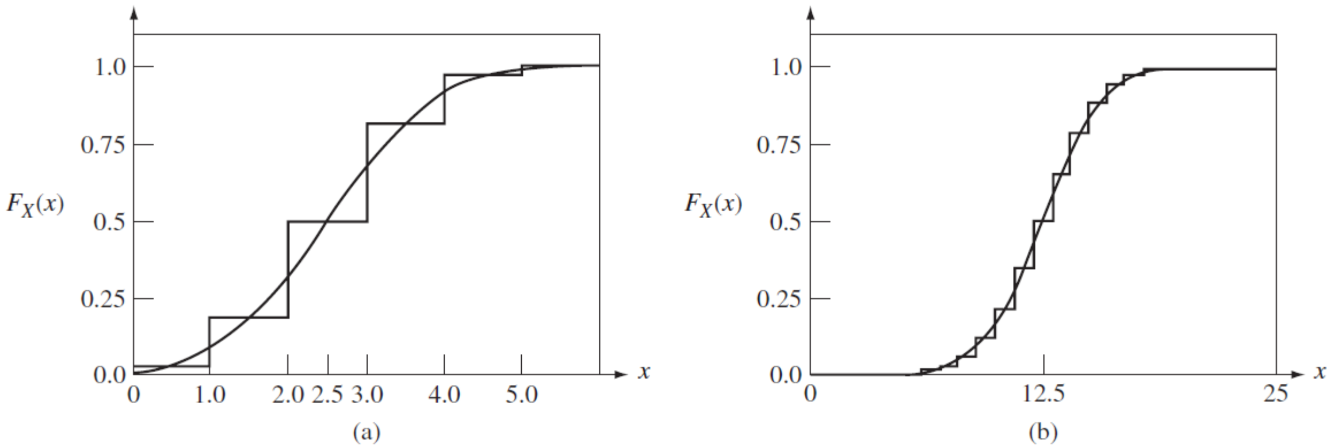
3. For large n, treat Z_n as if standard normal (Gaussian (0,1))

$$F_{Z_n}(z) \approx \Phi(z)$$

$$F_{w_n}(w) = F_{Z_n}(\frac{w - n\mu}{\sqrt{n\sigma^2}}) \approx \Phi(\frac{w - n\mu}{\sqrt{n\sigma^2}})$$

Treat W_n as if Gaussian $(n\mu, \sqrt{n\sigma^2})$.

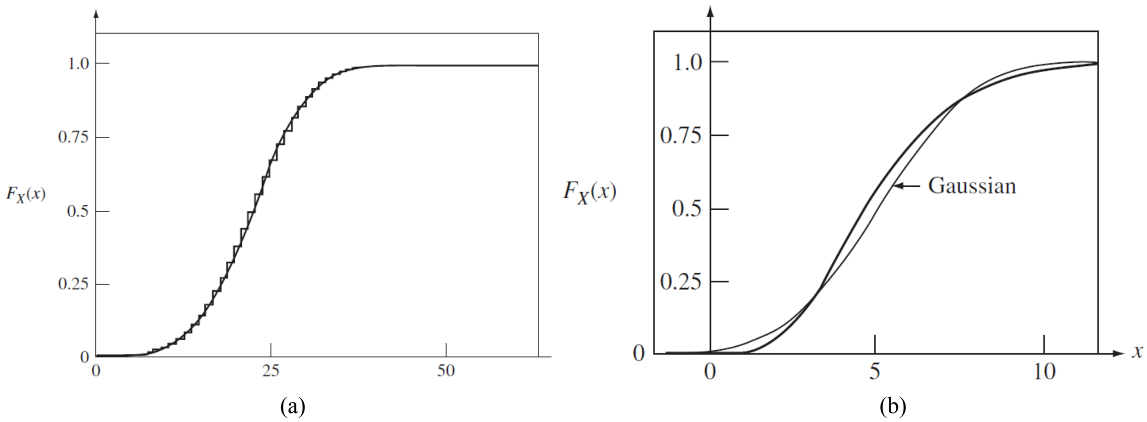
Example 1



(a) The CDF of the sum of 5 independent Bernoulli RVs and the CDF of a Gaussian RV of the same mean and variance.

(b) The CDF of the sum of 25 independent Bernoulli RVs and the CDF of a Gaussian random variable of the same mean and variance.

Example 2



(a) The CDF of the sum of 5 independent discrete, uniform RVs from the set $\{0, 1, \ldots, 9\}$ and the CDF of a Gaussian random variable of the same mean and variance .

(b) The CDF of the sum of 5 independent exponential RVs of mean 1 and the CDF of a Gaussian random variable of the same mean and variance.

Central Limit Theorem

- The Central limit theorem is very general, the only conditions are:
  - independence of X_is
  - finite mean and variance
- The distribution of X_is can be continuous, discrete or mixed
- The Central limit theorem provides a justification for why Gaussian random variables occur so often in natural and man-made phenomena: Many macroscopic phenomena result from the addition of numerous independent microscopic elements. Often times we are interested in averages consisting of a sum of independent random variables.

Example

The access times X to get one block of information from a computer disk are independent of one another and are uniformly distributed between 0 and 12 milliseconds. Before performing a certain task, the computer must access 12 different blocks of information from the disk. The total access time for all the information is a random variable A milliseconds.

1) What is the expected value of access time X?
2) What is the variance of access time X?
3) What is the expected value of the total access time A?
4) What is the standard deviation of the total access time A?
5) Use the central limit theorem to estimate the probability that the total access time exceeds 75 ms.
6) Use the central limit theorem to estimate the probability that the total access time is less than 48 ms.

1. $E[X] = \frac{1}{2}(0+12) = \boxed{6 \text{ ms}}$

2. $Var[X] = \frac{1}{12}(12-0)^2 = \boxed{12}$

3. $E[A] = nE[X] = 12(6) = \boxed{72 \text{ ms}}$

4. $\sigma[A] = \sqrt{n Var[X]} = \sqrt{12 \cdot 12} = \boxed{12}$

5. $P[A > 75] = 1 - P[A \leq 75]$

$$= 1 - P\left[Z \leq \frac{75-72}{12}\right]$$

$$= 1 - P[Z \leq 0.25]$$

$$= Q\left(\frac{1}{4}\right) = \boxed{0.4}$$

6. $P[A < 48] = P\left[Z \leq \frac{48-72}{12}\right]$

$$= P[Z \leq -2]$$

$$= Q(2) = \boxed{0.0228}$$

Stochastic Processes

- In some random experiments, the outcome of the experiment is a function of time or space:
  - Temporally varying: temperature of your body, electrical noise at output of equipment
  - Spatially varying: roughness of sand on the ground
  - Both temporally and spatially varying: waves on a pond, atmospheric temperature, ocean temperature

Stochastic Processes

- In some random experiments, the outcome of the experiment is a function of time or space:
  - Temporally varying: temperature of your body, electrical noise at output of equipment
  - Spatially varying: roughness of sand on the ground
  - Both temporally and spatially varying: waves on a pond, atmospheric temperature, ocean temperature
- More examples of time-varying outcomes:
  - The number of customers in a queuing system varies with time
  - The sequence of daily prices of a stock market
  - The sequence of scores in a football game
  - The traffic load in a communication network
  - The radar measurements of the position of an airplane

Definition

- A *random* or *stochastic process* $X(t)$ consists of an experiment with a probability measure $P[\cdot]$ defined on a sample space S and a function that assigns a **time function** $X(t, s)$ to each outcome s in the sample space of the experiment.

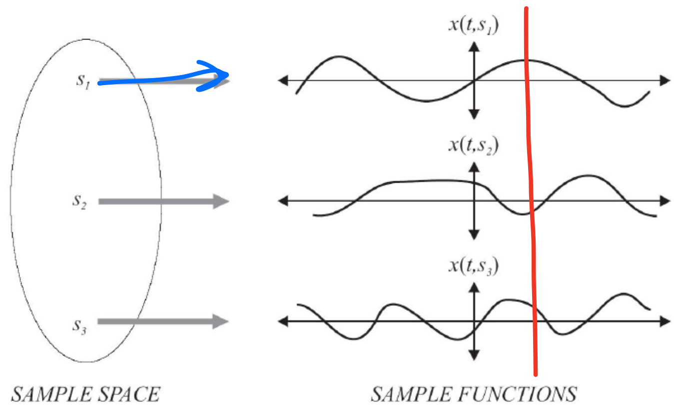
Definition

- A *random* or *stochastic process* $X(t)$ consists of an experiment with a probability measure $P[\cdot]$ defined on a sample space S and a function that assigns a **time function** $X(t, s)$ to each outcome s in the sample space of the experiment.

Note: Recall the definition of a random variable: A random variable X consists of an experiment with a probability measure $P[\cdot]$, defined on a sample space S and a function that assigns a *real number* to each outcome in S.

- Just as a random variable assigns a *number* to each outcome s in the sample space, a stochastic process assigns a *sample function* to each outcome s.

Stochastic Processes

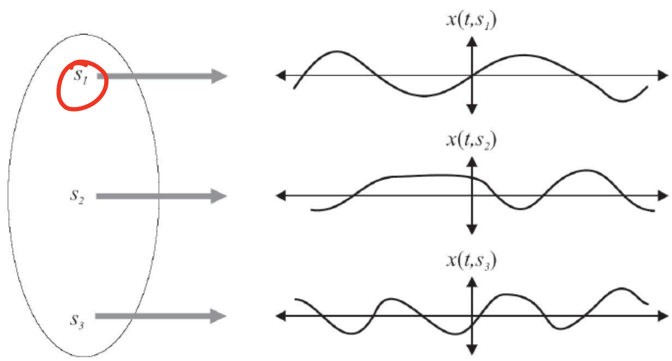


- Fix t – looks like a regular RV
- Fix s – just get a function

2 Degrees of randomness

- Notation:
  - $X(t)$ denotes the stochastic process
  - s denotes the particular outcome of the experiment
  - t indicates time dependence

Stochastic Processes



- Notation:
  - $X(t)$ denotes the stochastic process
  - s denotes the particular outcome of the experiment
  - t indicates time dependence
- The function $X(t, s)$, for fixed s, is called a **sample function**.
- Set of all possible time functions $X(t, s)$ that can result from an experiment is called the **ensemble** of a stochastic process.

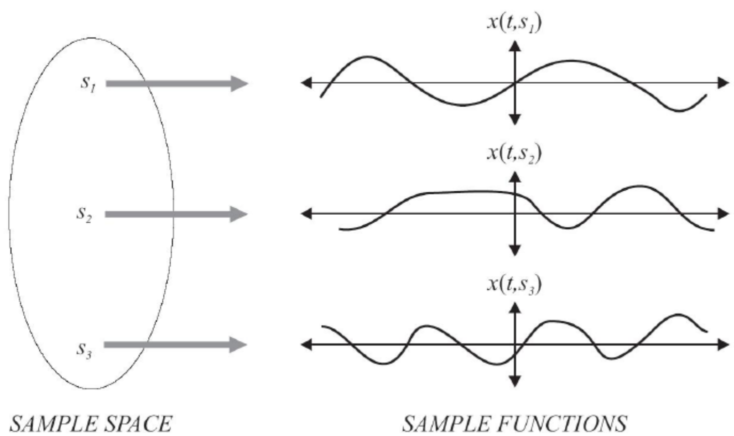
Example: Single toss of a fair coin

- We toss a fair coin once, the outcome of the experiment includes the events $S = \{H,\ T\}$. We associate the following time functions with the outcome of the experiment:

$$\begin{cases} X(t, H) = \sin t & \text{for } s = H \\ X(t, T) = 5 & \text{for } s = T \end{cases}$$

- The probability for each of the two outcomes is $\frac{1}{2}$.
- The collection of these two functions $X(t, H)$ and $X(t, T)$ and their corresponding probabilities provide a complete description of the stochastic process, where the set $\{\sin t, 5\}$ is the ensemble of the stochastic process.

Stochastic Processes



- A stochastic process is a function of both s and t:
  - If s is fixed at $s = s_0$ and t is variable then $X(t, s_0)$ is a time function.
  - If t is fixed at $t = t_0$ and s is variable then $X(t_0, s)$ is a random variable.
  - If s and t are both fixed at $s = s_0$ and $t = t_0$ then $X(t_0, s_0)$ is a real number.

Averages

- Since a stochastic process is a function of two variables t and s, there are two kinds of averages associated with it:
  - **Ensemble Average**: When t is fixed at $t = t_0$, $X(t_0)$ is a random variable and the *ensemble average* in this case corresponds to the expected value of that random variables.
  - **Time Average**: If s is fixed at $s = s_0$, then the sample function $X(t, s_0)$ is a function of time and we can define a *time average* of the sample function.

Averages

- Since a stochastic process is a function of two variables t and s, there are two kinds of averages associated with it:
  - **Ensemble Average**: When t is fixed at $t = t_0$, $X(t_0)$ is a random variable and the *ensemble average* in this case corresponds to the expected value of that random variables.
  - **Time Average**: If s is fixed at $s = s_0$, then the sample function $X(t, s_0)$ is a function of time and we can define a *time average* of the sample function.

Example The noontime temperature at Pittsburgh airport is measured daily for one year in consecutive years: the sequence $T(1)$, $T(2)$,..., $T(365)$ represents the temperature measurements.

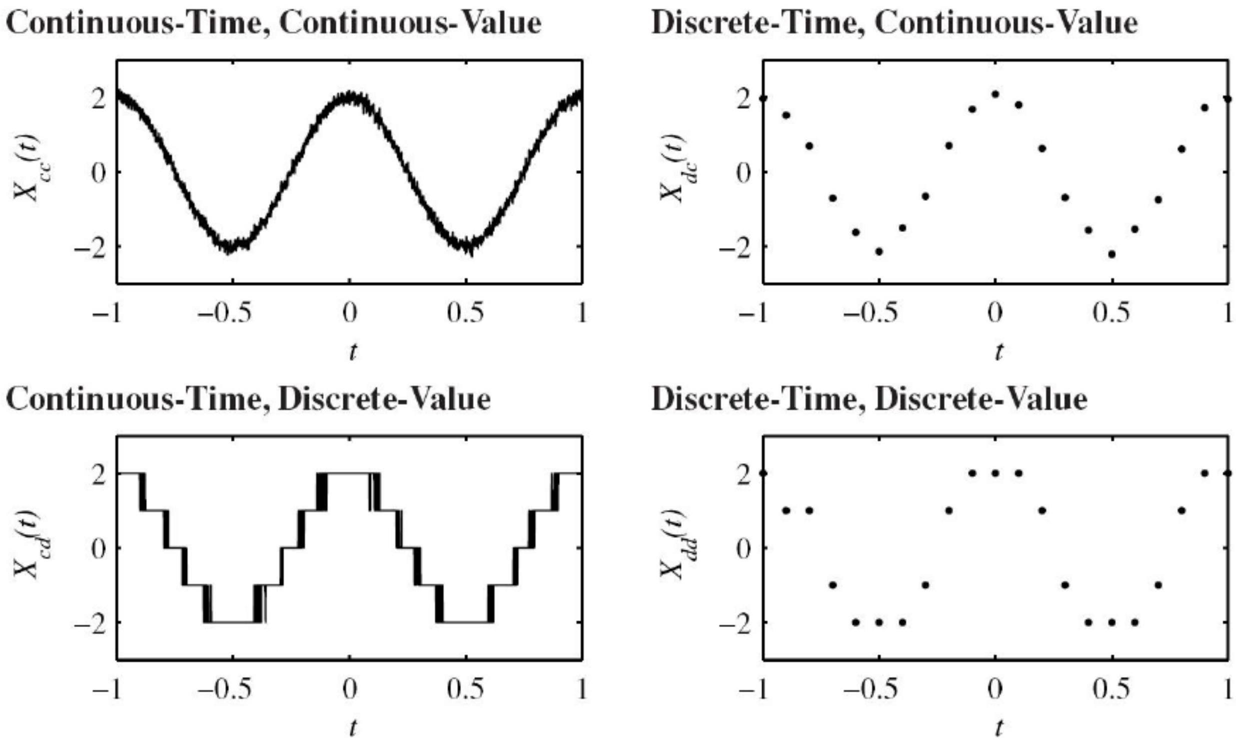
- The ensemble average corresponds to "the average noontime temperature for February 26" over all years for which it's measured
- The time average corresponds to "the average noontime temperature for 2017".

Types of Stochastic Processes

- **Discrete-Value vs. Continuous-Value Processes**
  The stochastic process $X(t)$ is a *discrete-value* process if the set S of all possible outcomes of $X(t)$ is a countable set; otherwise $X(t)$ is a *continuous-value* stochastic process.

- **Discrete-Time vs. Continuous-Time Processes**
  The stochastic process $X(t)$ is a *discrete-time* process if $X(t)$ is defined for a countable set of time instances I where $t \in I$; otherwise $X(t)$ is a *continuous-time* stochastic process.

Types of Stochastic Processes



Example 1

- Let s be a number selected at random from the interval $S = [0, 1)$ and let $b_1, b_2, \ldots$ be the binary expansion of s:

$$s = \sum_{i=1}^{\infty} b_i 2^{-i} \quad \text{where } b_i \in \{0, 1\}$$

$\frac{b_1}{2} + \frac{b_2}{4} + \cdots$

Define the discrete-time random process $X(n)$ by

$$X(n) = b_n \quad n = 1, 2, \ldots$$

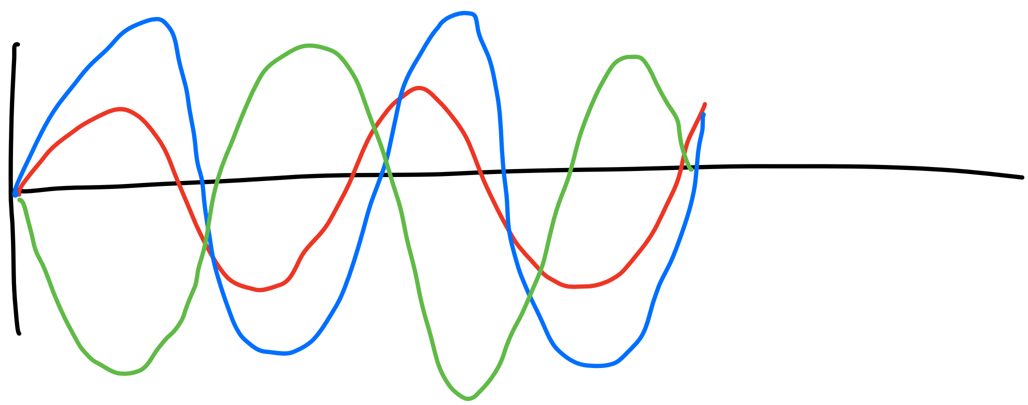
- The resulting process is a sequence of binary numbers, with $X(n)$ equal to the nth number in the binary expansion.

Example 2

- Let s be a number selected at random from the interval $S = [-1, 1]$. Define the continuous-time stochastic process $X(t)$ with the following sample functions: Continuous value

$$X(t, s) = s\cos(2\pi t) \quad -\infty \leq t \leq \infty$$

- The sample functions of this stochastic process are sinusoidal with amplitude $s \in [-1, 1]$ as shown below in Figure (a).



Example 2

- Let s be a number selected at random from the interval $S = [-1, 1]$. Define the continuous-time stochastic process $X(t)$ with the following sample functions:

$$X(t, s) = s\cos(2\pi t) \quad -\infty \le t \le \infty$$

- The sample functions of this stochastic process are sinusoidal with amplitude $s \in [-1, 1]$ as shown below in Figure (a).
- Let s be selected uniformly at random from interval $(-\pi, \pi)$ and define the continuous-time stochastic process $Y(t)$ with the following sample functions: Discrete value

$$Y(t, s) = \cos(2\pi t + s)$$

- The sample function of the stochastic process $Y(t)$ are time-shifted versions of $\cos(2\pi t)$ as shown below in Figure (b).

Example 2 (continued)