

I225E Statistical Signal Processing

4. Stochastic Process and Systems I


MAWALIM and UNOKI

candylin@jaist.ac.jp and unoki@jaist.ac.jp

School of Information Science

Agenda

- Review
- Practice Exercise Discussion
- Introduction to Stochastic System
- Memoryless System



REVIEW OF PROBABILITY THEORY & BASICS OF STOCHASTIC PROCESS

Review Terminology

■ Stochastic Process

Definition: A stochastic process is a family of random variables,

$$\{X(t) : t \in T\},$$

where t usually denotes time.

■ Discrete-time Process

Definition: if the set T is finite or countable.

In practice, this generally means $T = \{0, 1, 2, 3, \dots\}$

■ Continuous-time Process

Definition: if T is not finite or countable.

In practice, this generally means $T = [0, \infty)$, or $T = [0, K]$ for some K .

Review Terminology

■ State Space

Definition: The state space, S , is the set of real values that $X(t)$ can take.

Every $X(t)$ takes a value in \mathbb{R} , but S will often be a smaller set: $S \subseteq \mathbb{R}$.

For example, if $X(t)$ is the outcome of a coin tossed at time t , then the state space is $S = \{0, 1\}$.

■ Discrete and Continuous

The state space S is *discrete* if it is finite or countable. Otherwise, it is *continuous*.

Review Terminology

■ Independent Random Variables

Consider two discrete random variables X and Y . We say that X and Y are independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y), \text{ for all } x, y.$$

Remember, the concept of independent events A and B .

$$P(A, B) = P(A \cap B) = P(A)P(B)$$

Review Terminology

■ Expected value (=mean=average)

$$EX = E[X] = E(X) = \mu_X$$

■ Theorem:

■ $E[aX + b] = aE[X] + b$, for all $a, b \in \mathbb{R}$

■ $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$, for any set of random variables X_1, X_2, \dots, X_n .

■ Expected value of a function of a random variable:

$$E[g(X)] = \sum_{x_k \in R_X} g(x_k) P_X(x_k)$$

Review Terminology

■ Variance

$$\text{Var}(X) = E[(X - \mu_X)^2].$$

$$\blacksquare \text{Var}(X) = E[X^2] - [EX]^2$$

■ Standard Deviation

$$SD(X) = \sigma_X = \sqrt{\text{Var}(X)}.$$

■ Theorem:

■ For a random variable X and real numbers a and b , $\text{Var}(aX + b) = a^2 \text{Var}(X)$

■ If X_1, X_2, \dots, X_n are independent random variables and $X = X_1 + X_2 + \dots + X_n$, then

$$\text{Var}(X) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

Special Distribution (Discrete)

■ Bernouli distribution

■ **Notation:** $X \sim \text{Bernouli}(p)$.

■ **Description:** A random variable X is a Bernoulli random variable with parameter p , where $0 < p < 1$.

■ **Probability function:**

$$f_X(x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

■ **Mean:** $E(X) = p$.

■ **Variance:** $\text{Var}(X) = pq$, where $q = 1 - p$.

Special Distribution (Discrete)

■ Binomial distribution

■ **Notation:** $X \sim \text{Binomial}(n, p)$.

■ **Description:** number of successes in n independent trials, each with probability p of success.

■ **Probability function:**

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \text{ for } x = 0, 1, \dots, n.$$

■ **Mean:** $E(X) = np$.

■ **Variance:** $\text{Var}(X) = np(1 - p) = npq$, where $q = 1 - p$.

■ **Sum:** If $X \sim \text{Binomial}(n, p)$, $Y \sim \text{Binomial}(m, p)$, and X and Y are independent, then

$$X + Y \sim \text{Binomial}(n + m, p).$$

Special Distribution (Discrete)

■ Poisson distribution

■ **Notation:** $X \sim \text{Poisson}(\lambda)$.

■ **Probability function:**

$$f_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \text{ for } x = 0, 1, 2, \dots$$

■ **Mean:** $E(X) = \lambda$.

■ **Variance:** $\text{Var}(X) = \lambda$.

■ **Sum:** If $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$, and X and Y are independent, then

$$X + Y \sim \text{Poisson}(\lambda + \mu).$$

Special Distribution (Continuous)

■ Uniform distribution

■ **Notation:** $X \sim \text{Uniform}(a, b)$.

■ **Probability density function (PDF):**

$$f_X(x) = \frac{1}{b-a}, \text{ for } a < x < b$$

■ **Cumulative distribution function:**

$$F_X(x) = P(X \leq x) = \frac{x-a}{b-a}, \text{ for } a < x < b$$

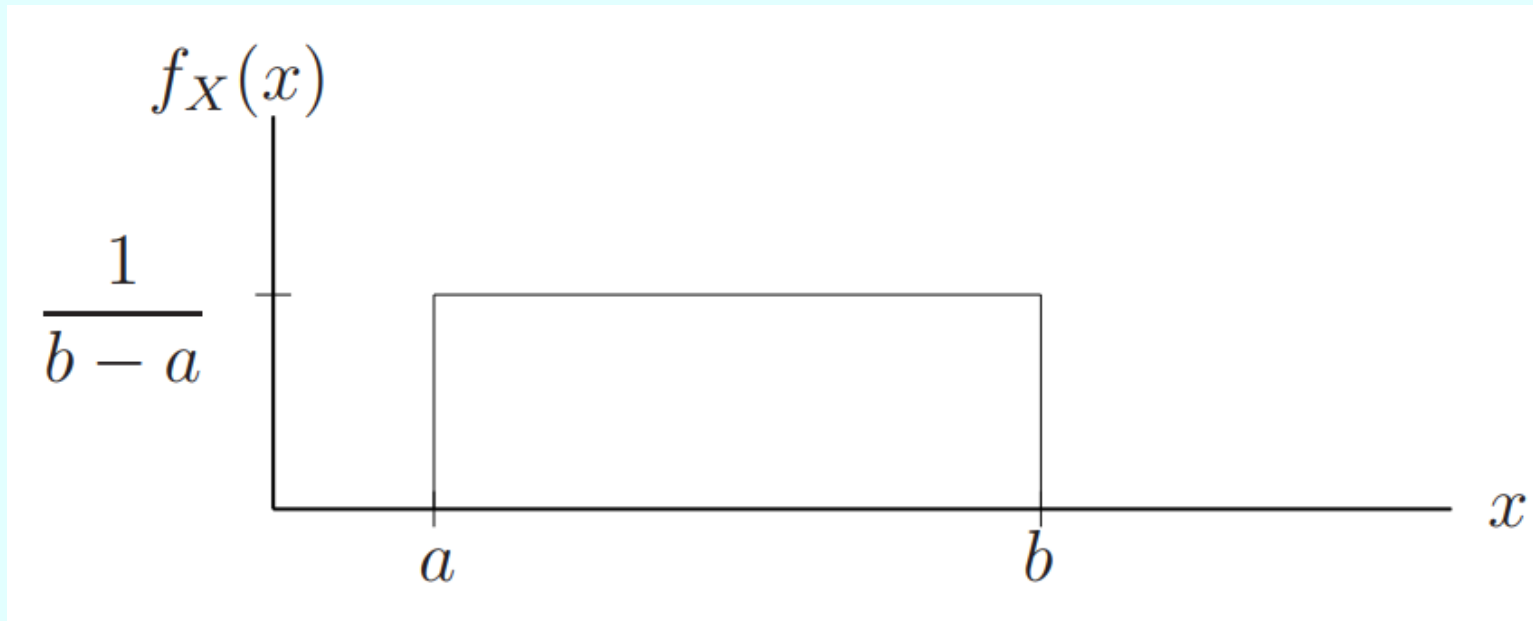
$$F_X(x) = 0, \text{ for } x \leq a, \text{ and } F_X(x) = 1, x \geq b$$

■ **Mean:** $E(x) = \frac{a+b}{2}$

■ **Variance:** $\text{Var}(X) = \frac{(b-a)^2}{12}$

Special Distribution (Continuous)

■ Uniform distribution (a, b)



Special Distribution (Continuous)

■ Exponential distribution

■ **Notation:** $X \sim \text{Exponential}(\lambda)$.

■ **Probability density function (PDF):**

$$f_X(x) = \lambda e^{-\lambda x}, \text{ for } 0 < x < \infty$$

■ **Cumulative distribution function:**

$$F_X(x) = P(X \leq x) = 1 - e^{-\lambda x}, \text{ for } 0 < x < \infty$$

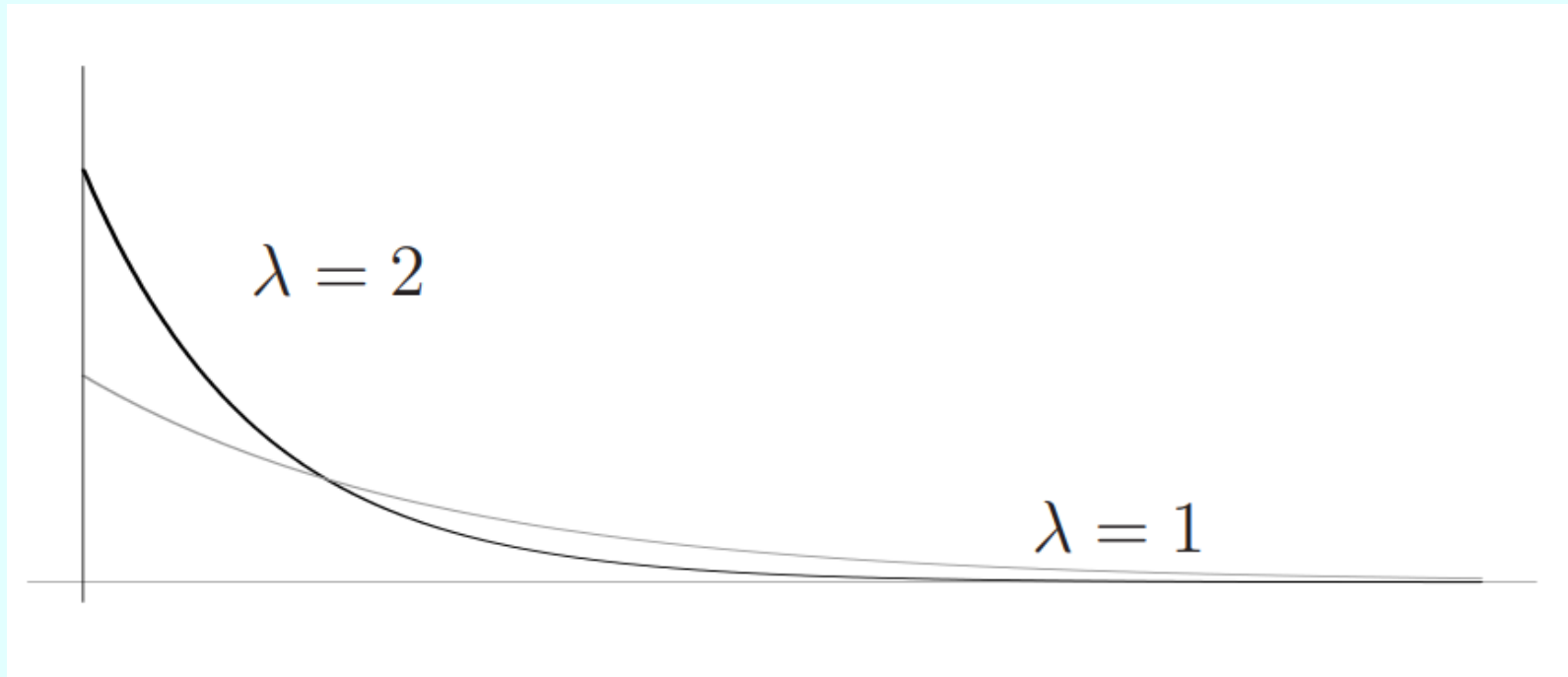
$$F_X(x) = 0, \text{ for } x \leq 0$$

■ **Mean:** $E(x) = \frac{1}{\lambda}$

■ **Variance:** $\text{Var}(X) = \frac{1}{\lambda^2}$

Special Distribution (Continuous)

■ Exponential distribution (λ)



Special Distribution (Continuous)

■ Gamma distribution

■ **Notation:** $X \sim \text{Gamma}(k, \lambda)$.

■ **Probability density function (PDF):**

$$f_X(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}, \text{ for } 0 < x < \infty$$

where $\Gamma(k) = \int_0^\infty y^{k-1} e^{-y} dy$ (the Gamma function).

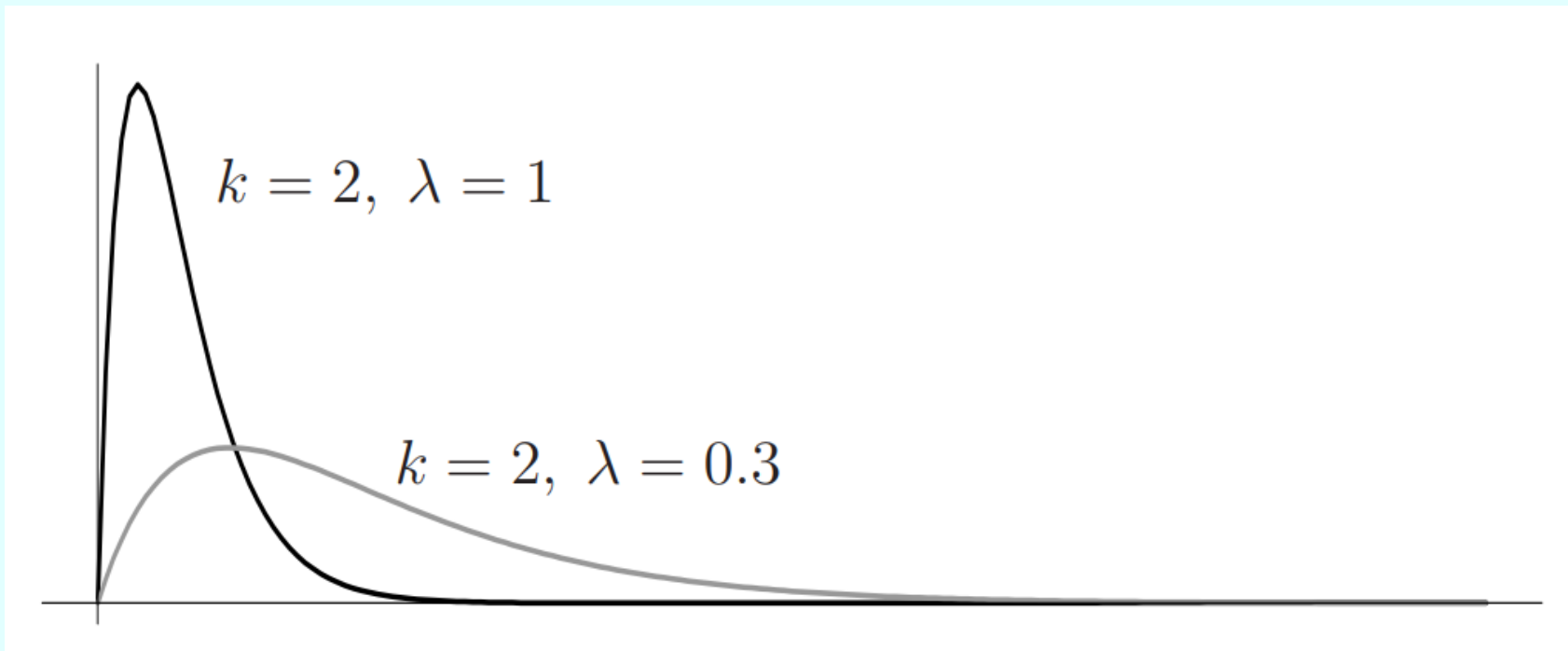
■ **Cumulative distribution function:** no closed form.

■ **Mean:** $E(x) = \frac{k}{\lambda}$

■ **Variance:** $\text{Var}(X) = \frac{k}{\lambda^2}$

Special Distribution (Continuous)

■ Gamma distribution (k, λ)



Special Distribution (Continuous)

■ Normal distribution

■ **Notation:** $X \sim \text{Normal}(\mu, \sigma^2)$.

■ **Probability density function (PDF):**

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}, \text{ for } -\infty < x < \infty$$

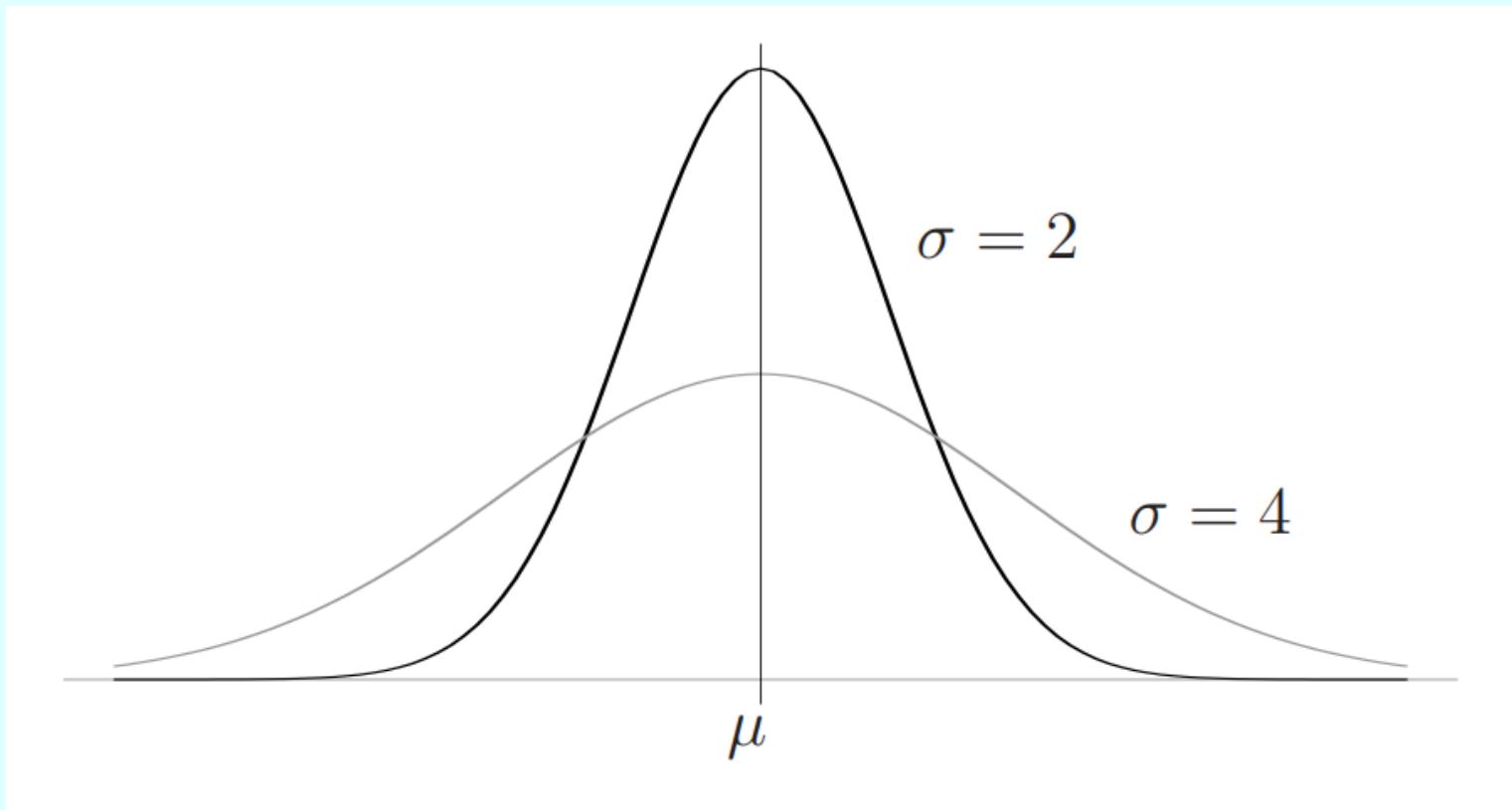
■ **Cumulative distribution function:** no closed form.

■ **Mean:** $E(x) = \mu$

■ **Variance:** $\text{Var}(X) = \sigma^2$

Special Distribution (Continuous)

■ Normal distribution (μ, σ^2)



Autocorrelation Function (ACF)

■ Autocorrelation of $X(t)$

$$\begin{aligned}\rho_X(t_1, t_2) &= \text{Corr}(X_{t_1}, X_{t_2}) = \\ R_{XX}(t_1, t_2) &= E\{X(t_1)X(t_2)\} = \\ &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_X(x_1, x_2; t_1, t_2) dx_1 dx_2\end{aligned}$$

where **sample autocorrelation** is

$$\bar{R}_{XX}(t_1, t_2) = \frac{1}{n} \sum_{t=1}^n X(t + t_1)X(t + t_2)$$

Covariance Function

- The autocovariance function $\gamma_X(t_1, t_2)$ of a stochastic process $\{X_t\}$ measures the covariance between the process at two different time points t_1 and t_2 :

$$\begin{aligned}\gamma_X(t_1, t_2) &= \text{Cov}(X_{t_1}, X_{t_2}) = C_{XX}(t_1, t_2) \\ &= E \left[\left(X_{t_1} - \mu_X(t_1) \right) \left(X_{t_2} - \mu_X(t_2) \right) \right] \\ &= R_{XX}(t_1, t_2) - \eta_X(t_1)\eta_X(t_2)\end{aligned}$$

- For a weakly stationary process, the autocovariance function depends only on the time difference (lag) $k = t_2 - t_1$.
- In the case of $t_1 = t_2 = t$, $C_{XX}(t_1, t_2)$ is equal to variance of
$$X(t) \rightarrow C_{XX}(t, t) = E\{X(t)X(t)\} - \eta_X^2(t) = \text{Var}(X(t))$$

■ Complex process

$X(t) = Y(t) + jZ(t)$: complex variable $X(t)$ is composed of real part $Y(t)$ and imaginary part $Z(t)$.

$$\text{Corr}(X_{t_1}, X_{t_2}) = R_{XX}(t_1, t_2) = E\{X(t_1)X^*(t_2)\}$$

$$\text{Corr}(X_{t_1}, X_{t_2}) = R_{XX}(t_2, t_1) = R_{XX}^*(t_1, t_2)$$

$$\text{Corr}(X_t, X_t) = R_{XX}(t, t) = E\{|X(t)|^2\} \geq 0$$

$$\text{Cov}(X_{t_1}, X_{t_2}) = C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \eta_X(t_1)\eta_X^*(t_2)$$

■ Correlation coefficient

$$r(t_1, t_2) = \frac{C_{XX}(t_1, t_2)}{\sqrt{C_{XX}(t_1, t_1)C_{XX}(t_2, t_2)}}$$

Previous Exercise

[Exercise]

A complex stochastic process $Z(t)$ is given by:

$$Z(t) = X(t) + iY(t)$$

Where:

$X(t)$ and $Y(t)$ are real-valued random processes.

i is the imaginary unit.

$$E[X(t)] = 1, E[Y(t)] = 2, \text{ for all } t.$$

$$\text{Var}[X(t)] = 4, \text{Var}[Y(t)] = 9, \text{ for all } t.$$

$$\text{Cov}(X(t), Y(t)) = 0, \text{ for all } t.$$

The autocorrelation function of $X(t)$ is $R_{XX}(t_1, t_2) = 4 \cdot \exp\left(-\frac{|t_1 - t_2|}{2}\right)$.

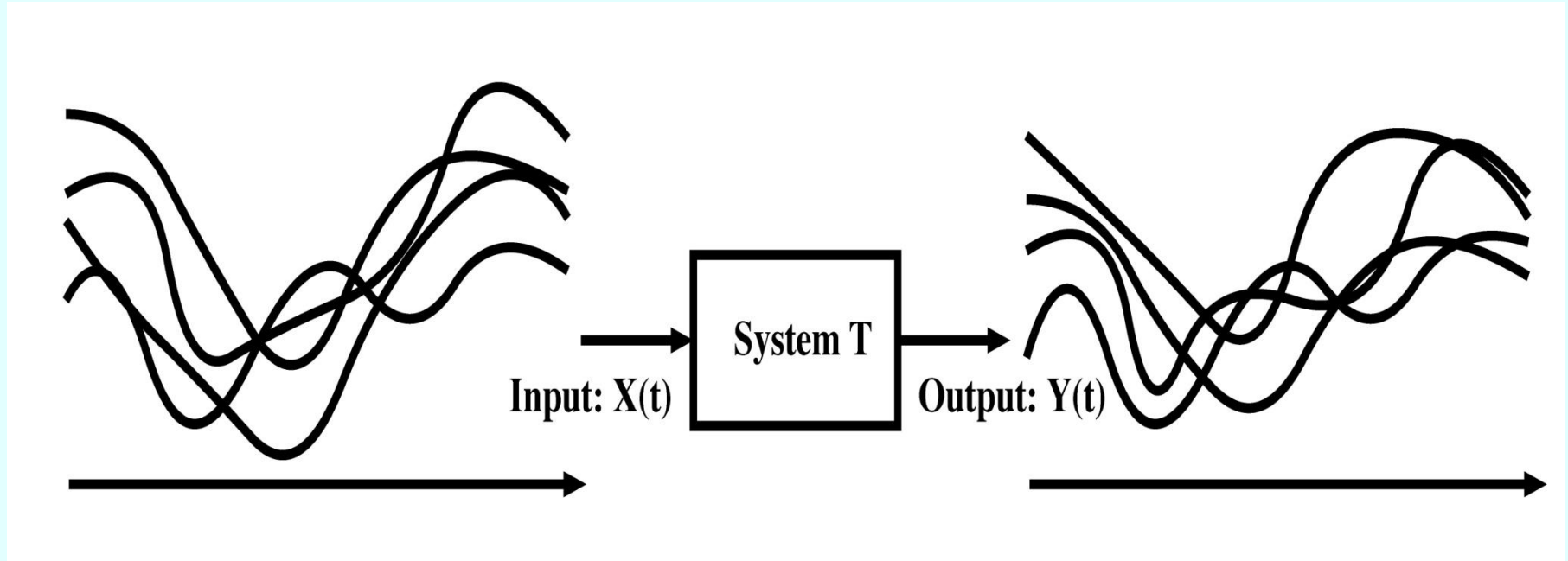
The autocorrelation function of $Y(t)$ is $R_{YY}(t_1, t_2) = 9 \cdot \cos(\pi * (t_1 - t_2))$.

Previous Exercise

[Exercise]

- a) Calculate the mean function of the complex process $Z(t)$.
- b) Calculate the autocovariance function of the complex process $Z(t)$.
- c) Calculate the autocorrelation function of the complex process $Z(t)$.
- d) Calculate the correlation coefficient between $Z(t_1)$ and $Z(t_2)$.

1. System with stochastic input



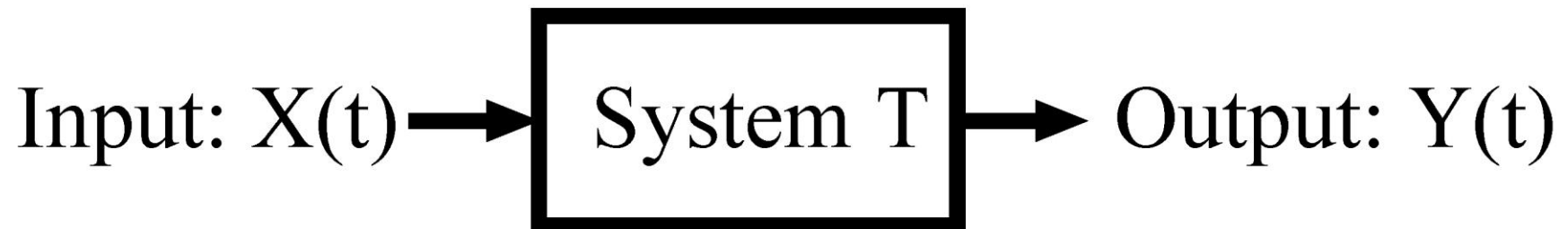
■ System

Given a stochastic process $X(t)$ as input, $Y(t)$ represents its output.

$$Y(t) = T[X(t)]$$

■ Purpose:

If statistical properties of input $X(t)$ are known, study the statistical properties of output $Y(t)$



$$Y(t) = T[X(t)]$$

System Dynamics

■ Deterministic system:

System operates only on variable t , treating outcome ω as a parameter. Namely,

If $X(t, \omega_1) = X(t, \omega_2)$, then $Y(t, \omega_1) = Y(t, \omega_2)$,

■ Stochastic system:

System operates on both t and ω . Namely,

Even if $X(t, \omega_1) = X(t, \omega_2)$, $Y(t, \omega_1) \neq Y(t, \omega_2)$.

Example: Physical element of the system or coefficient of the system equation is stochastic.



$$Y(t) = T[X(t)]$$

- This lecture deals with only **deterministic systems**.
- In deterministic systems, transformation T may depend on t . To emphasize this, sometimes denoted as

$$Y(t) = T_t[X(t)]$$

referred to as a time-dependent system.

Deterministic System

Memoryless System

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

System with Memory

**Time-Varying
System**

**Time-Invariant
System**

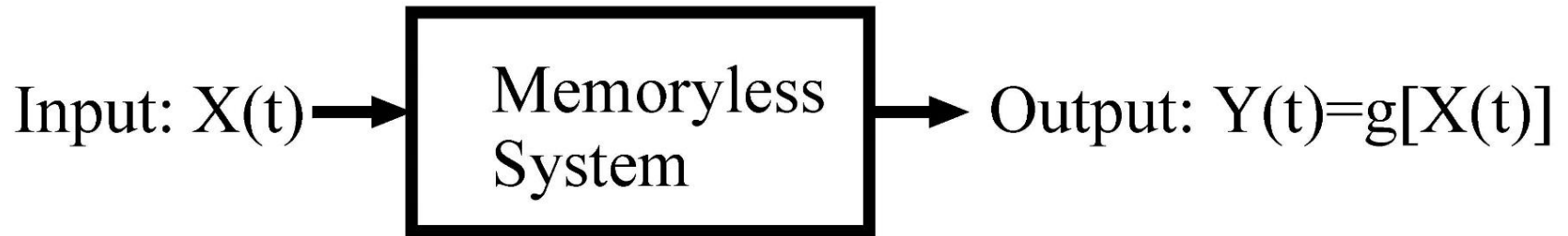
Linear System

$$Y(t) = L[X(t)]$$

Linear Time-Invariant system

$$Y(t) = X(t) * h(t)$$

2. Memoryless System



- **Output** $Y(t_1)$ at time $t = t_1$ depends only upon the simultaneous state of input $X(t_1)$, but not upon past or future state of $X(t)$

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

Mean of the Output

- The mean (or expected value) of the output process $Y(t)$, denoted by $E[Y(t)]$ or $\mu_Y(t)$, can be found using the law of the unconscious statistician:

$$(a1) \mu_Y(t) = E\{Y(t)\} = E[g(X(t))] = \int_{-\infty}^{\infty} g(x) f_X(x, t) dx$$

where $f_X(x, t)$ is the first-order probability density function (PDF) of the input process $X(t)$ at time t .

Correlation of the Output

- The autocorrelation function of the output process $Y(t)$, denoted by $R_Y(t_1, t_2) = E[Y(t_1)Y(t_2)]$, measures the statistical dependence between the output at two different times t_1 and t_2 .
- For a memoryless system, $Y(t_1)$ depends only on $X(t_1)$, and $Y(t_2)$ depends only on $X(t_2)$. Therefore, the autocorrelation of the output is:

$$\begin{aligned} \text{(a2)} \quad E\{Y(t_1)Y(t_2)\} &= E\{g(X(t_1))g(X(t_2))\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1)g(x_2)f_X(x_1, x_2; t_1, t_2)dx_1dx_2 \end{aligned}$$

(a3) n th-order density of $Y(t)$, $f_Y(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n)$ is obtained from n th-order density of $X(t)$,

$f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)$ through transformation

$$Y(t_1) = g[X(t_1)], Y(t_2) = g[X(t_2)], \dots, Y(t_n) = g[X(t_n)].$$

If the following system $y_1 = g[x_1], y_2 = g[x_2], \dots, y_n = g[x_n]$ has a unique solution $x = [x_1, x_2, \dots, x_n]$, n th-order density of $Y(t)$ is obtained as

$$f_Y(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n) = \frac{f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{|J(x_1, x_2, \dots, x_n)|}$$

where J is Jacobian $J = [g'(x_1)g'(x_2) \dots g'(x_n)]$.

When more than two solutions exist, summation of the corresponding terms $\frac{f_X}{|J|}$ gives the n th-order density.

Digression on coordinate transformation

- Let us consider n -dim variables $\mathbf{x} = (x_1, \dots, x_n)^T$ and $\mathbf{y} = (y_1, \dots, y_n)^T$ and the mapping $y_i = y_i(x_1, \dots, x_n) = y_i(\mathbf{x})$. In this case, their infinitesimal volume are related as

$$dy_1 \cdots dy_n = |J(x_1, \dots, x_n)| dx_1 \cdots dx_n$$

where the matrix is called the Jacobian:

$$J = \begin{pmatrix} \partial y_1 / \partial x_1 & \cdots & \partial y_1 / \partial x_n \\ \vdots & \ddots & \vdots \\ \partial y_n / \partial x_1 & \cdots & \partial y_n / \partial x_n \end{pmatrix}$$

- Accordingly, their probability densities are related as

$$f_{\mathbf{y}}(y_1, \dots, y_n) = \frac{1}{|J(x_1, \dots, x_n)|} f_{\mathbf{x}}(x_1, \dots, x_n)$$

Appendix

Following has been used for the derivation of the density in (a3).

With respect to random variables $\mathbf{X} = [X_1, X_2, \dots, X_n]$, n functions

$$Y_1 = g_1(\mathbf{X}), Y_2 = g_2(\mathbf{X}), \dots, Y_n = g_n(\mathbf{X}),$$

are given. For n random numbers $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]$, we determine their joint density $f_Y(y_1, y_2, \dots, y_n)$, where y_1, y_2, \dots, y_n represent a specific set of numbers.

To find the density, we solve the system

$$g_1(\mathbf{X}) = y_1, g_2(\mathbf{X}) = y_2, \dots, g_n(\mathbf{X}) = y_n.$$

If the system has no solution, then $f_Y(y_1, y_2, \dots, y_n) = 0$. If the system has a single solution $\mathbf{x} = [x_1, x_2, \dots, x_n]$, the density can be obtained by substituting the solution into following formula

$$f_Y(y_1, y_2, \dots, y_n) = \frac{f_X(x_1, x_2, \dots, x_n)}{|J(x_1, x_2, \dots, x_n)|},$$

where

$$J(x_1, x_2, \dots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

If more than two solutions exist, the density is given by summation of all the corresponding terms

$$f_Y = \left. \frac{f_X}{|J|} \right|_{\mathbf{X}=\mathbf{X}_1} + \left. \frac{f_X}{|J|} \right|_{\mathbf{X}=\mathbf{X}_2} + \dots$$

[Exercise]

Imagine you have a transformation that takes a point (u, v) in the uv -plane and maps it to a point (x, y) in the xy -plane. Let's say this transformation is defined by the following equations:


$$\begin{aligned}x(u, v) &= u^2 v \\ y(u, v) &= u + v^2\end{aligned}$$

Calculate the Jacobian matrix of the output variables $(x$ and $y)$ with respect to the input variables $(u$ and $v)$.



[Answer]

$$J(u, v) = \begin{pmatrix} 2uv & u^2 \\ 1 & 2v \end{pmatrix}$$



(a4) If input $X(t)$ is strict sense stationary, output $Y(t)$ is also strict sense stationary.

[Proof] According to (a3), n th-order density of $Y(t)$ is given as

$$f_Y(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n) = \frac{f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{|J(x_1, x_2, \dots, x_n)|}$$

Since $X(t)$ is strict sense stationary, its density is invariant to a shift of the origin in time, denominator is independent of t . Therefore, f_Y is also invariant to time-shift. This proves that $Y(t)$ is strict sense stationary.



■ From the properties of strict sense stationary

(i) First-order density of $Y(t)$ is independent of t
 $\rightarrow f_Y(y; t) = f_Y(y)$

(ii) Second-order density is a function of time lag $\tau = t_1 - t_2$
 $\rightarrow f_Y(y_1, y_2; t_1, t_2) = f_Y(y_1, y_2; \tau)$

Example of memoryless system

■ Square-law detector

Square-law detector is a memoryless system whose output equals

$$Y(t) = X^2(t).$$

Using the density $f_X(x, t)$ of input $X(t)$, find the density $f_Y(y, t)$ of output $Y(t)$.

■ First-order density

If $y > 0$, solutions of $y = x^2$ are $x = \pm\sqrt{y}$.

The corresponding Jacobian matrices are $J = \frac{dx^2}{dx} = 2x = \pm 2\sqrt{y}$.

Hence

$$\begin{aligned} f_Y(y; t) &= \left. \frac{f_X}{|J|} \right|_{x=\sqrt{y}} + \left. \frac{f_X}{|J|} \right|_{x=-\sqrt{y}} \\ &= \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}; t) + f_X(-\sqrt{y}; t)] \end{aligned}$$

■ Second-order density:

If $y_1 > 0$, $y_2 > 0$, solutions of $y_1 = x_1^2$, $y_2 = x_2^2$ are $(\pm\sqrt{y_1}, \pm\sqrt{y_2})$. Since the corresponding Jacobian matrices are $J = \pm 4\sqrt{y_1 y_2}$,

$$f_Y(y_1, y_2; t_1, t_2) = \frac{1}{4\sqrt{y_1 y_2}} \sum f_X(\pm\sqrt{y_1}, \pm\sqrt{y_2}; t_1, t_2)$$