Chapter 2: **Review** of Probability Theory

2.1. Probability

2.2. Random variable

2.3. Function of random variable

2.4. Discrete random vector

2.5. Moments of a Discrete Random Variable

2.6. Examples of Random variables

2.7. Sum of random variable

2.8. Random process

2.1. Probability

2.1.1. Probability definition

2.1.2. Conditional probability

2.1.3. Bayes rule

2.1.4.Total Probability theorem

2.1.5. Independence between events

2.1.1. Probability definition

- Use *Measure Theory*
- Based on a triplet

$$(\Omega, F, P)$$

Where:

- Ω: sample space
  - Set of all possible outcomes
- F: σ−algebra
  - Set of all possible events or combinations of outcomes
- P: probability function
  - Any set function
    - Domain is Ω
    - Range is the closed unit interval $[0,1]$

2.1.1. Probability definition (Cont.)

- P must obey the following rules:
  - $P(\Omega) = 1$
  - Let A be any event in F, then $P(A) \geq 0$
  - Let A and B be two events in F such that $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.
  - Probability of complement: $P(A) = 1 - P(A)$.
  - $P(A) \leq 1$.
  - $P(\emptyset) = 0$.
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2.1.2. Conditional probability (Cont.)

- Let A and B be two events, with $P(A) > 0$. The conditional probability of B given A is defined as:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- Hence, $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$
- If $A \cap B = \emptyset$ then $P(B|A) = 0$
- If $A \subset B$, then $P(B|A) = 1$

2.1.3.Bayes rule

- If A and B are events:

$$P(A|B) = \frac{P(B/A)P(A)}{P(B)}$$

2.1.4.Total Probability Theorem

- A set of B_i, $i = 1, \ldots, n$ of events is a partition of Ω when:
  - $\bigcup_{i=1}^{n} B_i = \Omega$.
  - $B_i \cap B_j = \emptyset$, if $i \neq j$.
- Theorem: If A is an event and B_i, $i = 1, \ldots, n$ of is a partition of Ω, then:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

2.1.5. Independence between Events

- Two events A and B are statistically independent when

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

- Supposing that both $P(A)$ and $P(B)$ are greater than zero, from the above definition we have that:

$$P(A|B) = P(A) \qquad P(B|A) = P(B)$$

2.1.5. Independence between Events (Cont.)

- N events are statistically independent if the intersection of the events contained in any subset of those N events have probability equal to the product of the individual probabilities
- Example: Three events A, B and C are independent if:

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

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

2.2. Random variable

2.1.1. What is random variables?

2.1.2. Cumulative Distribution Function (CDF)

2.2.3. Types of random variable

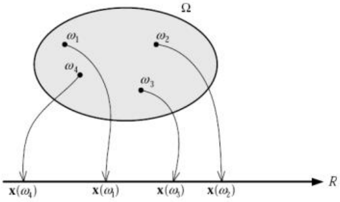
2.2.4. Probability Density Function (PDF)

2.2.1. What is random variables?

- A random variable (rv) is a function that maps each $\omega \in \Omega$ to a real number

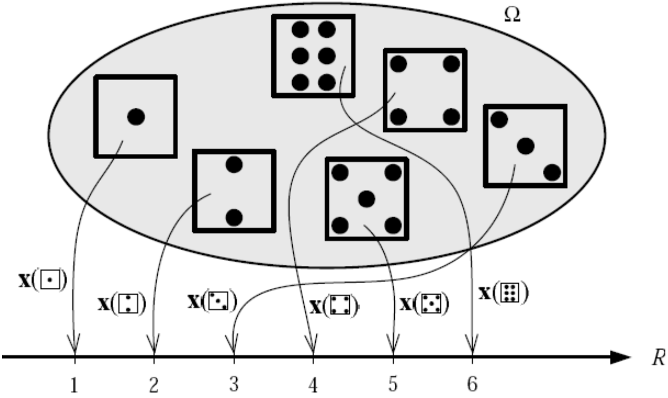
$$X \;:\; \Omega \to \mathrm{R}$$
$$\omega \to X(\omega)$$

E.g:



- A random variable is a *mapping* from the sample space Ω to the set of real numbers.

2.2.1. What is random variables? (cont.)



2.2.1. What is random variables? (cont.)

- Random variable is a representation in numeric form of sample space. Thus,
  - We can process events with digital computing devices
  - Values corresponding to the outcomes are sorted

2.2.1. What is random variables? (cont.)

- Through a random variable, subsets of Ω (w) are mapped as interval (I) of the real numbers. Hence:

$$P(X \in I) = P(\{w | X(\omega) \in I\})$$

Probability of the interval

Probability of the subset

2.2.1. What is random variables? (cont.)

Ex: Throw a dice with 6 faces

- Maps the sample space (1 dot, 2 dots…6 dots) to numbers (1,2…6).
    - The probability of each overcome is 1/6. The probability of each number is also 1/6
- The subset (1 dot, 2 dots, 3 dots) maps to interval from 1 to 3.
    - The probability of the subset will be 3/6. The probability of the interval will be also 3/6

2.2.2. Cumulative Distribution Function (CDF)

- CDF of a random variable X with a given value x ($F_X(x)$) is the probability for a random variable X has a value that does not exceed x

$$F_X(x) = P(X \leq x)$$

- $F_X(x)$ (∞) = 1
- $F_X(x)$ (-∞) = 0
- If $x_1 < x_2, F_X(x_2) \geq F_X(x_1)$

2.2.3. Types of random variable

- Discrete: Cumulative function is a step function (sum of unit step functions)

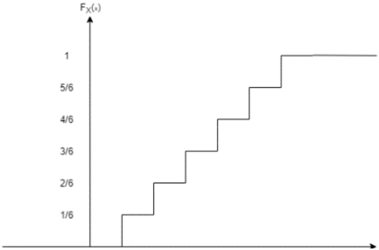
$$F_X(x)=\sum_i P\ (X = x_i)\ u\ (x - x_i)$$

where $u(x)$ is unit step function

2.2.3. Types of random variable (Cont.)

Example: X is the random variable that describes the outcome of the roll of a die $X \in \{1, 2, 3, 4, 5, 6\}$

P(x)=1/6 for all x



- Continous: Cumulative function is a continous function.
- Mixed: Neither discrete nor continous.

2.2.4. Probability Density Function (PDF)

- It is the derivative of the cumulative distribution function:

$$p_X(x) = \frac{d}{dx} F_X(x)$$

- $\int_{-\infty}^{x} p_X(x)dx = F_X(x)$.
- $p_X(x) \geq 0$.
- $\int_{-\infty}^{\infty} p_X(x)dx = 1$.
- $\int_{a}^{b} p_X(x)dx = F_X(b) - F_X(a) = P(a \leq X \leq b)$.

2.2.4. Probability Density Function (PDF) (Cont.)

- Discrete random variable:
  - P_X(x) = P(X=x)
  - The sum taken for every value of x of the probability density function must be equal to 1

$$\sum_{X} px(x) = 1$$

2.3. Function of random variable

- Random variable X = {x}, $x_{min} \leq$ x $\leq x_{min}$ with F_X(x)
- Y = G(X) is defined as function of random variable
  - Generate random variable Y ={y}, (y_{min} = G(x_{min}) $\leq$ y $\leq y_{max}$ = G(x_{max}))
  - Need to determine F_Y(y) from G(X) and F_X(x)
  - Example: Y = aX + b, a and b are constants, a > 0

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) =$$

$$\int_{-\infty}^{\frac{y-b}{a}} p_X(x)dx = F_X(\frac{y-b}{a})$$

2.3. Function of random variable (Cont.)

- Generalization:

$$p_Y(y) = \sum_{i=1}^{n} \frac{p_X(x_i)}{|g'(x_i)|}$$

Where x_i is the solution of equation g(x) = y

- Coding operation is to map a source to a new source which is suitable with transmission requirements
  - E.g. source coding is to map an information source to a code source that has uniform probability distribution
  - From random variable function, a mapping can be found to transfer from an arbitrary source to a source that has required probability distribution

2.4. Discrete Random vector

Let $Z = [X, Y]$ be a random vector with sample space $Z = X \times Y$
X and Y are random variables have sample space

The joint probability distribution function of Z is mapping

$$p_Z(z) : \mathcal{Z} \rightarrow [0,1] \text{ satisfying:}$$

$$\sum_{Z \in \mathcal{Z}} p_Z(z) = \sum_{x,y \times \mathcal{Y}} p_{XY}(x,y) = 1$$

2.4. Discrete Random vector (Cont.)

- Marginal Distributions:

$$p_X(x) = \sum_{y \in Y} p_{XY}(x, y)$$

$$p_Y(y) = \sum_{x \in X} p_{XY}(x, y)$$

2.4. Discrete Random vector (Cont.)

- Conditional Distributions:

$$p_{X|Y=y}(\mathrm{x}) = \frac{p_{XY}(x,y)}{p_Y(y)}$$

$$p_{Y|X=x}(\mathrm{y}) = \frac{p_{XY}(x,y)}{p_X(x)}$$

2.4. Discrete Random vector (Cont.)

- Random variables X and Y are independent if and only if

$$p_{XY}(x, y) = p_X(x)p_Y(y)$$

- Consequences:

$$p_{X|Y=y}(x) = p_X(x)$$

$$p_{Y|X=x}(y) = p_Y(y)$$

2.5. Moments of a Discrete Random Variable

- The $n-$th order moment of a discrete random variable X is defined as:

$$E[X^n] = \sum_{x \in \mathcal{X}} x^n p_X(x)$$

- if $n = 1$, we have the mean of X, $m_X = E[X]$.
- The $m-$th order central moment of a discrete random variable X is defined as:

$$E[(X - m_X)^m] = \sum_{x \in \mathcal{X}} (x - m_X)^m p_X(x)$$

- if $m = 2$, we have the variance of X, σ_X^2.

2.5. Moments of a Discrete Random Variable

- The joint moment $n-$th order with relation to X and $k-$th order with relation to Y:

$$m_{nk} = E[X^n Y^k] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} x^n y^k p_{XY}(x, y)$$

- The joint central $n-$th order with relation to X and $k-$th order with relation to Y:

$$\mu_{nk} = E[(X - m_X)^n (Y - m_Y)^k] == \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (x - m_X)^n (y - m_Y)^k p_{XY}(x, y)$$

2.5. Moments of a Discrete Random Variable

- The correlation of two random variables X and Y is the expected value of their product (joint moment of order 1 in X and order 1 in Y):

$$Corr(X, Y) = m_{11} = E[XY]$$

- The covariance of two random variables X and Y is the joint central moment of order 1 in X and order 1 in Y:

$$Cov(X, Y) = \mu_{11} = E[(X - m_X)(Y - m_Y)]$$

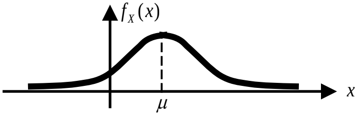
- $Cov(X, Y) = Corr(X, Y) - m_X m_Y$
- Correlation Coefficient:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad \rightarrow \quad -1 \leq \rho_{XY} \leq 1$$

2.6. Examples of Random variables

- Normal (Gaussian) random variables
  - Density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$



  - Bell shape curve
  - Distribution function

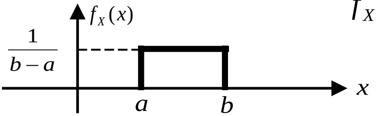
$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy = G\left(\frac{x-\mu}{\sigma}\right),$$

$$G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

2.6. Examples of Random variables (Cont.)

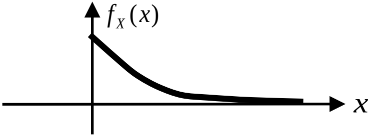
- Uniform random variables: $X \sim U(a,b),\ a < b,$
  - Density function

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$



- Exponential random variables $X \sim \varepsilon(\lambda)$
  - Density function

$$f_X(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$



2.7. Sum of Gaussian random variable

- Gaussian independent random variables x_1, x_2,..., x_n have mean m_1, m_2,… m_n, variance σ_{x_1}, σ_2,…,σ_{x_n}
- Sum of all variables y =$\sum_1^n x_i$ have
  - Mean m = $\sum_1^n m_i$: constant one-dimensional component
  - Variance σ = $\sum_1^n \sigma_i$: Average power (alternating current power)
- E.g. A channel has input X, noise N, output Y = X+N
  - Noise and Input: independent
  - m_Y = m_X + m_N
  - P_Y = P_X + P_N with P(.) is average power

2.8. Random process

2.8.1 . Definition

2.8.2. Stationary random process

2.8.3. Statistical average or joint moments

2.8.4. Mean value or the first moment

2.8.5. Mean-squared value or the second moment

2.8.6. Correlation

2.8.7. Power spectral density of a random process

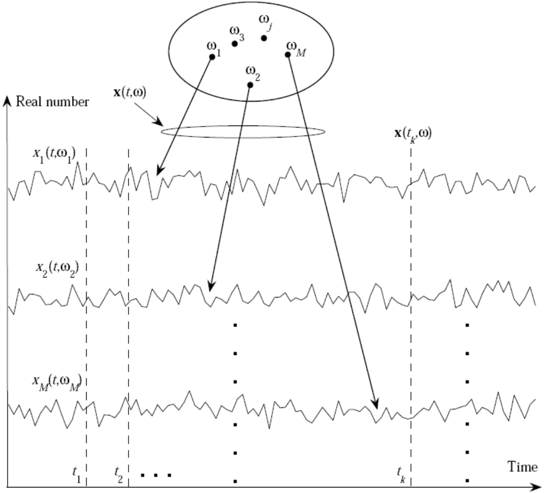
2.8.8. Time averaging and Egordicity

2.8.9. Random process and LTI system

2.8.1. Definition

- Random process is a set of potential random variables (realization, member) to describe a physical object
- Example:
  - To describe the movements of an animal, it is necessary to have different cameras
    - Sample space of each camera is a random variable

2.8.1. Definition (Cont.)



A mapping from a sample space to *a set of time functions.*

2.8.1 Definition (Cont.)

- To describe the random process, we need to have a set of values
- It is necessary to have the probability of appearing simultaneously at one time on a time function of the realization

$$f_{\mathbf{x}(t_1),\mathbf{x}(t_2)}(x_1, x_2; t_1, t_2)$$

2.8.2. Stationary random process

- Based on whether its statistics change with time: the process is *non-stationary* or *stationary*.
- Different levels of stationarity:
  - Strictly stationary: the joint pdf of any order is independent of a shift in time.
  - Nth-order stationarity: the joint pdf does not depend on the time shift

$$f_{\mathbf{x}(t_1),\mathbf{x}(t_2),\ldots\mathbf{x}(t_N)}(x_1, x_2, \ldots, x_N; t_1, t_2, \ldots, t_N) =$$
$$f_{\mathbf{x}(t_1+t),\mathbf{x}(t_2+t),\ldots\mathbf{x}(t_N+t)}(x_1, x_2, \ldots, x_N; t_1 + t, t_2 + t, \ldots, t_N + t).$$

- The first- and second-order stationarity:

$$f_{\mathbf{x}(t_1)}(x, t_1) = f_{\mathbf{x}(t_1+t)}(x; t_1 + t) = f_{\mathbf{x}(t)}(x)$$

$$f_{\mathbf{x}(t_1),\mathbf{x}(t_2)}(x_1, x_2; t_1, t_2) = f_{\mathbf{x}(t_1+t),\mathbf{x}(t_2+t)}(x_1, x_2; t_1 + t, t_2 + t)$$
$$= f_{\mathbf{x}(t_1),\mathbf{x}(t_2)}(x_1, x_2; \tau), \quad \tau = t_2 - t_1.$$

2.8.3. Statistical average or joint moments

- Consider N random variables $\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots \mathbf{x}(t_N)$. The joint moments of these random variables is

$$E\{\mathbf{x}^{k_1}(t_1), \mathbf{x}^{k_2}(t_2), \ldots \mathbf{x}^{k_N}(t_N)\} = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_N=-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \cdots x_N^{k_N} f_{\mathbf{x}(t_1),\mathbf{x}(t_2),\ldots\mathbf{x}(t_N)}(x_1, x_2, \ldots, x_N; t_1, t_2, \ldots, t_N) \, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_N,$$

  for all integers $k_j \geq 1$ and $N \geq 1$.
- Shall only consider the first- and second-order moments, i.e., $E\{\mathbf{x}(t)\}$, $E\{\mathbf{x}^2(t)\}$ and $E\{\mathbf{x}(t_1)\mathbf{x}(t_2)\}$. They are the *mean value*, *mean-squared value* and *(auto)correlation*.

2.8.4. Mean value or the first moment

- The mean value of the process at time t is

$$m_{\mathbf{x}}(t) = E\{\mathbf{x}(t)\} = \int_{-\infty}^{\infty} x f_{\mathbf{x}(t)}(x;t)\mathrm{d}x.$$

- The average is *across the ensemble* and if the pdf varies with time then the mean value is a (deterministic) function of time.
- If the process is stationary then the mean is independent of t or a constant:

$$m_{\mathbf{x}} = E\{\mathbf{x}(t)\} = \int_{-\infty}^{\infty} x f_{\mathbf{x}}(x)\mathrm{d}x.$$

2.8.5. Mean-squared value or the second moment

- This is defined as

$$\mathsf{MSV}_{\mathbf{x}}(t) = E\{\mathbf{x}^2(t)\} = \int_{-\infty}^{\infty} x^2 f_{\mathbf{x}(t)}(x;t)\mathrm{d}x \text{ (non-stationary)},$$

$$\mathsf{MSV}_{\mathbf{x}} = E\{\mathbf{x}^2(t)\} = \int_{-\infty}^{\infty} x^2 f_{\mathbf{x}}(x)\mathrm{d}x \text{ (stationary)}.$$

- The second central moment (or the variance) is:

$$\begin{aligned}
\sigma_{\mathbf{x}}^2(t) &= E\left\{[\mathbf{x}(t) - m_{\mathbf{x}}(t)]^2\right\} = \mathsf{MSV}_{\mathbf{x}}(t) - m_{\mathbf{x}}^2(t) \text{ (non-stationary)}, \\
\sigma_{\mathbf{x}}^2 &= E\left\{[\mathbf{x}(t) - m_{\mathbf{x}}]^2\right\} = \mathsf{MSV}_{\mathbf{x}} - m_{\mathbf{x}}^2 \text{ (stationary)}.
\end{aligned}$$

2.8.6. Correlation

- The autocorrelation function completely describes the power spectral density of the random process.
- Defined as the correlation between the two random variables $\mathbf{x}_1 = \mathbf{x}(t_1)$ and $\mathbf{x}_2 = \mathbf{x}(t_2)$:

$$\begin{aligned} R_{\mathbf{x}}(t_1, t_2) &= E\{\mathbf{x}(t_1)\mathbf{x}(t_2)\} \\ &= \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} x_1 x_2 f_{\mathbf{x}_1,\mathbf{x}_2}(x_1, x_2; t_1, t_2) \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

- For a stationary process:

$$\begin{aligned} R_{\mathbf{x}}(\tau) &= E\{\mathbf{x}(t)\mathbf{x}(t+\tau)\} \\ &= \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} x_1 x_2 f_{\mathbf{x}_1,\mathbf{x}_2}(x_1, x_2; \tau) \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

- *Wide-sense stationarity* (WSS) process: $E\{\mathbf{x}(t)\} = m_{\mathbf{x}}$ for any t, and $R_{\mathbf{x}}(t_1, t_2) = R_{\mathbf{x}}(\tau)$ for $\tau = t_2 - t_1$.

2.8.7. Power spectral density of a random process

- Taking the Fourier transform of the random process does not work
  - Each representation has a unique spectrum
- Need to determine how the average power of the process is distributed in frequency.

- It can be shown that *the power spectral density and the autocorrelation function are a Fourier transform pair*:

$$R_{\mathbf{x}}(\tau) \longleftrightarrow S_{\mathbf{x}}(f) = \int_{\tau=-\infty}^{\infty} R_{\mathbf{x}}(\tau) \mathrm{e}^{-j2\pi f\tau} \mathrm{d}\tau.$$

2.8.8. Time averaging and Egordicity

- Time averaging:

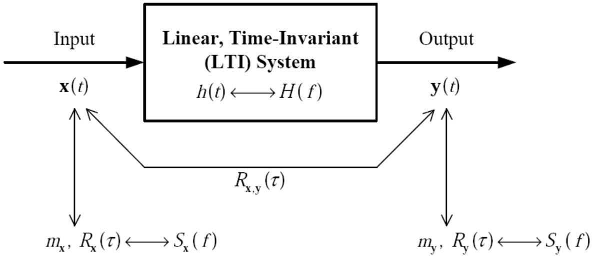
All time averages on a single ensemble member are equal to the corresponding ensemble average:

$$
\begin{aligned}
E\{\mathbf{x}^n(t))\} &= \int_{-\infty}^{\infty} x^n f_{\mathbf{x}}(x)\mathrm{d}x \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} [\mathbf{x}_k(t,\omega_k)]^n \mathrm{d}t, \ \forall\ n,\ k.
\end{aligned}
$$

- Ergodicity: must be stationary and time averaging = Statistical average

For an ergodic process: To measure various statistical averages, it is sufficient to look at only one realization of the process and find the corresponding time average.

2.8.9. Random process and LTI system



$$
\begin{aligned}
m_{\mathbf{y}} &= E\{\mathbf{y}(t)\} = E\left\{\int_{-\infty}^{\infty} h(\lambda)\mathbf{x}(t-\lambda)\mathrm{d}\lambda\right\} = m_{\mathbf{x}} H(0) \\
S_{\mathbf{y}}(f) &= |H(f)|^2 S_{\mathbf{x}}(f) \\
R_{\mathbf{y}}(\tau) &= h(\tau) * h(-\tau) * R_{\mathbf{x}}(\tau).
\end{aligned}
$$