ECE801
Monitoring and Estimation

Background

Instructor: Christos Panayiotou

Outline

- Vectors and Matrices
- Probability and Random Variable
- Stochastic Processes

Vectors and Matrices

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \Re^n \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \in \Re^{n \times m}$$

Vector

Matrix

Function of multiples variables.

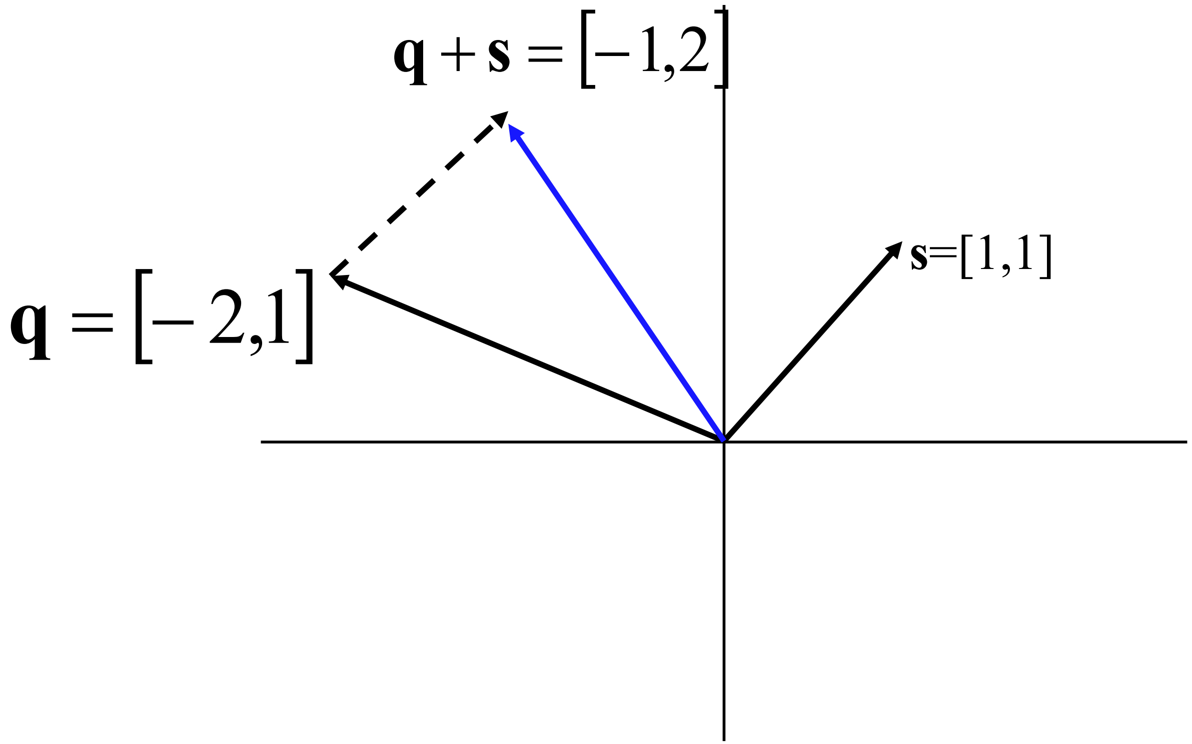
$$f(x_1, \cdots, x_n) = f(\mathbf{x})$$

Vector and Matrix Addition

If $\mathbf{x} \in \Re^n$ and $\mathbf{y} \in \Re^n$ $\quad \mathbf{x} \pm \mathbf{y} = \begin{bmatrix} x_1 \pm y_1 \\ x_2 \pm y_2 \\ \vdots \\ x_n \pm y_n \end{bmatrix}$

If $\mathbf{A}, \mathbf{B} \in \Re^{n \times m}$ $\quad \mathbf{A} \pm \mathbf{B} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1m} \pm b_{1m} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2m} \pm b_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} \pm b_{n1} & a_{n2} \pm b_{n2} & \cdots & a_{nm} \pm b_{nm} \end{bmatrix}$

Vector Addition

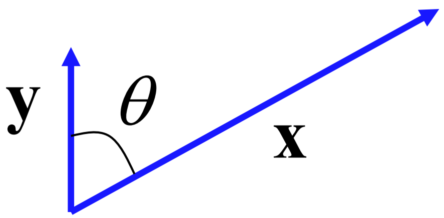


Inner (Dot) product

$$\mathbf{x}, \mathbf{y} \in \Re^n$$

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos\theta$$



Example

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathbf{x}^T \mathbf{y} = 1 \cdot 0 + 1 \cdot 1 = 1$$

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos\theta$$

$$= \sqrt{2} \cdot 1 \cdot \frac{\sqrt{2}}{2} = 1$$

$$\theta = 45^\circ$$

Matrix Multiplication

$\mathbf{A} \in \Re^{n \times m}, \mathbf{B} \in \Re^{m \times k}$

$$\mathbf{AB} = \begin{bmatrix} \sum_{l=1}^{m} a_{1l} b_{l1} & \cdots & \sum_{l=1}^{m} a_{1l} b_{lk} \\ \vdots & & \vdots \\ \sum_{l=1}^{m} a_{nl} b_{l1} & \cdots & \sum_{l=1}^{m} a_{nl} b_{lk} \end{bmatrix} \in \Re^{n \times k},$$

In general $\mathbf{AB} \neq \mathbf{BA}$

Properties

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$$

Vector and Matrix Transposition

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \Re^n \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \in \Re^{n\times m}$$

$$\mathbf{x}^T = \left[x_1, \cdots, x_n\right] \qquad \mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{mn} \end{bmatrix} \in \Re^{m\times n}$$

$$\left(\mathbf{AB}\right)^T = \mathbf{B}^T \mathbf{A}^T$$

Trace of a Matrix

- The trace of a square matrix $\boldsymbol{A}$ is the sum of its diagonal elements

$$trace[\boldsymbol{A}] = \sum_{i=1}^{n} a_{ii}$$

It also holds that

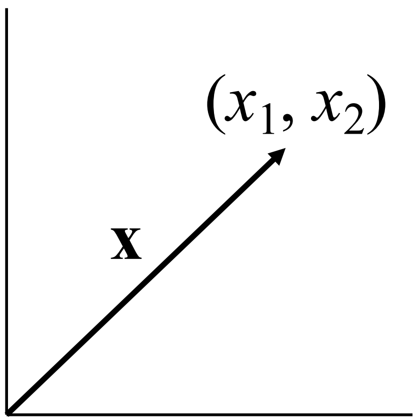
$$trace[\boldsymbol{AB}] = trace[\boldsymbol{BA}]$$

Rank of a Matrix

- The maximum number of linearly independent columns of a matrix
- A square $n \times n$ matrix is invertible if and only if it has full rank, i.e., if its rank is equal to n.

Vector Magnitude

$$d = \|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2}$$



In multi-dimensional spaces:

$$d(\mathbf{x}) = \|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Euclidean distance

Vector Functions

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}$$

Example:

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ 3x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \ln(x_1 + x_2 + x_3) \\ x_1 x_2 x_3 e^{x_1} \\ x_2 \end{bmatrix}$$

Gradient and Hessian

Partial Derivative:
$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{\delta \to 0} \frac{f(\mathbf{x} + \delta \mathbf{e}_i) - f(\mathbf{x})}{\delta}$$

Gradient
$$\mathbf{g} = \nabla f(\mathbf{x}) = \begin{bmatrix} \partial f(\mathbf{x}) / \partial x_1 \\ \vdots \\ \partial f(\mathbf{x}) / \partial x_n \end{bmatrix}$$

Hessian Matrix:
$$\mathbf{G} = \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \partial f^2(\mathbf{x}) / \partial x_1^2 & & \partial f^2(\mathbf{x}) / \partial x_1 \partial x_n \\ \partial f^2(\mathbf{x}) / \partial x_1 \partial x_2 & \ddots & \\ \vdots & & \\ \partial f^2(\mathbf{x}) / \partial x_1 \partial x_n & & \partial f^2(\mathbf{x}) / \partial x_n^2 \end{bmatrix}$$

Second derivatives
(curvature)

Examples

$$f(\mathbf{x}) = 100\left(x_2 - x_1^2\right)^2 + \left(1 - x_1\right)^2$$

$$\mathbf{g} = \nabla f(\mathbf{x}) = \begin{bmatrix} -400\left(x_2 - x_1^2\right)x_1 - 2\left(1 - x_1\right) \\ 200\left(x_2 - x_1^2\right) \end{bmatrix} \qquad \nabla f(\mathbf{0}) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$\mathbf{G} = \nabla^2 f(\mathbf{x}) = \begin{bmatrix} -400\left(x_2 - 3x_1^2\right) + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix} \quad \nabla^2 f(\mathbf{0}) = \begin{bmatrix} 2 & 0 \\ 0 & 200 \end{bmatrix}$$

$$f(\mathbf{x}) = x_1 \ln\left(x_1 + x_2\right)$$

$$\mathbf{g} = \nabla f(\mathbf{x}) = \begin{bmatrix} \ln\left(x_1 + x_2\right) + \dfrac{x_1}{x_1 + x_2} \\ \dfrac{x_1}{x_1 + x_2} \end{bmatrix} \qquad \mathbf{G} = \nabla^2 f(\mathbf{x}) = \frac{1}{\left(x_1 + x_2\right)^2} \begin{bmatrix} x_1 + 2x_2 & x_2 \\ x_2 & -1 \end{bmatrix}$$

Vector Function Derivatives

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix} \quad \nabla\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \partial f_1(\mathbf{x})/\partial x_1 & \cdots & \partial f_1(\mathbf{x})/\partial x_n \\ \vdots & & \vdots \\ \partial f_m(\mathbf{x})/\partial x_1 & \cdots & \partial f_m(\mathbf{x})/\partial x_n \end{bmatrix}$$

Jacobean Matrix

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} x_1 + x_2 + x_3 \\ 3x_1 x_2 e^{x_1 x_2} \end{bmatrix}$$

$$\nabla\mathbf{F}(\mathbf{x}) = \begin{bmatrix} 1 & 1 & 1 \\ 3x_2(1 + x_1 x_2)e^{x_1 x_2} & 3x_1(1 + x_1 x_2)e^{x_1 x_2} & 0 \end{bmatrix}$$

Determinant

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - cb$$

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \quad \det(\mathbf{A}) = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

$\boldsymbol{M}_{ij}$ is the matrix that results from $\mathbf{A}$ after we remove row i and column j.

$$\det[\boldsymbol{A}] = \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \det \boldsymbol{M}_{ij} \qquad \mathbf{A} \in \Re^{n \times n}$$

Eigenvalues and Eigenvectors

Definition: Assume that **A** has dimension nxn. We define as Eigenvalues the numbers λ for which

$$\mathbf{A}\mathbf{p} = \lambda\mathbf{p}$$

where $\mathbf{p}$ is a non-zero vector. The corresponding solutions $\mathbf{p}$ are the Eigenvectors of **A**.

Characteristic Polynomial:

$$f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$$

Degree n polynomial

The roots of the equation $f(\lambda)=0$ $(\lambda_1,\ldots, \lambda_n)$ are the Eigenvalues of A.

Eigenvalues and Eigenvectors

Example: Find the Eigenvalues and Eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 12 & 0 \\ 2 & 0 & 10 \end{bmatrix}$$

Characteristic Polynomial:

$$f(\lambda) = (12 - \lambda)^2 (8 - \lambda)$$ Eigenvalues= 12, 12, 8

Eigenvectors:

$$\mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i$$

$$\mathbf{p}_1 = \mathbf{p}_2 = r\begin{bmatrix} 1 & * & 1 \end{bmatrix}^T \quad \mathbf{p}_3 = r\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$$

Definite and Semidefinite Matrices

Assume $\mathbf{G}$ is an nxn symmetric matrix, then we define the quadratic function

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{G} \mathbf{x}$$

where $\mathbf{x}$ is a vector of dimension n. Then we say that

- $\mathbf{G}$ is positive definite if $Q(\mathbf{x})>0$ for all $\mathbf{x}\neq\mathbf{0}$.
- $\mathbf{G}$ is positive semidefinite if $Q(\mathbf{x})\geq 0$ for all $\mathbf{x}\neq\mathbf{0}$.
- $\mathbf{G}$ is negative definite if $Q(\mathbf{x})<0$ for all $\mathbf{x}\neq\mathbf{0}$.
- $\mathbf{G}$ is negative semidefinite if $Q(\mathbf{x})\leq 0$ for all $\mathbf{x}\neq\mathbf{0}$.

Probability

- Frequency definition of probability
  - Assume an experiment where there are n possible outcomes $A_1, A_2 \dots A_n$
  - Suppose we repeat the experiment k times and let N_i count the number of times we observe A_i, then

$$\Pr(A_i) = \lim_{k \to \infty} \frac{N_i}{k}$$

  - while it also holds

$$0 \leq \Pr(A_i) \leq 1$$

and

$$\sum_{i=1}^{n} \Pr(A_i) = 1$$

Joint outcomes/events

- If two outcomes A and B of the experiment are *independent*, then the probability of *both* occurring is

$$\Pr(AB) = \Pr(A)\Pr(B)$$

- If they are also *mutually exclusive*

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

while if they are not mutually exclusive

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$$

Conditional probability

- If two outcomes are not independent, then the occurrence of one outcome may tell us something about the occurrence of the other. Thus we define the *conditional probability* of the outcome A **given** that B has occurred

$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}$$

Which also implies that

$$\Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

and

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

Bayes' Rule

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

- Assume all possible mutually exclusive outcomes $A_1, A_2 \ldots A_n$, while B is some combination of these outcomes, then the law of total probability states that

$$\Pr(\mathrm{B}) = \sum_{i=1}^{n} \Pr(B|A_i)\Pr(A_i)$$

- And substituting in the Bayes' rule, above

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\sum_{i=1}^{n} \Pr(B|A_i)\Pr(A_i)}$$

Random Variables

- Random variables are mappings from the set of outcomes of a random experiment to the set of real numbers defined on a probability space
- Probability space $(\Omega, \mathcal{F}, P)$ where
  - Ω is the set of possible outcomes
  - $\mathcal{F}$ is the set of possible events where an even may consists from a set of possible outcomes (including the empty set)
  - P is the probability of an event
- Toss a coin with $\Omega = \{Heads, Tails\}$ and random variable $X(\omega); X(Heads) = 1; X(Tails) = 0.$
- Classification of random variables
  - Continuous random variables (take any real value)
  - Discrete random variables (take discrete (integer) values)

Distribution Functions

- (Cumulative) Distribution Function (cdf)

$$F_X(x) = \Pr[X \leq x] \text{ for all } x \in \mathbb{R}$$

  - $F_X(-\infty) = 0$
  - $F_X(\infty) = 1$
  - $F_X(x)$ is a non-decreasing function
- Joint distribution function

$$F_X(x_1, \dots, x_n) = \Pr[X_1 \leq x_1, \dots, X_n \leq x_n]$$

  - To obtain the marginal cdf $F(x_i)$ from the joint cdf use $x_j = \infty$ for all $j \neq i$.
- Independent random variables

$$F_X(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n)$$

Distribution Functions

- Probability Density Function (pdf) $f_X(x)$
  - Continuous variables

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$

  - Probability of the event $[a \leq X \leq b]$

$$\Pr[a \leq X \leq b] = F(b) - F(a) = \int_{a}^{b} f(y)dy$$

  - Note: $\Pr[X = x] = 0$
- Probability Mass Function
  - Discrete variables

$$F_X(x) = \sum_{y \leq x} \Pr[X = y]$$

Conditional Distributions

$$H(x, y) = \Pr[X \leq x | Y \leq y] = \frac{\Pr[X \leq x, Y \leq y]}{\Pr[Y \leq y]} = \frac{F(x,y)}{F(y)}$$

- What if the conditional event is $Y = y$, i.e., $\Pr[X \leq x | Y = y]$?
  - Define the conditional density function $f(x|y) = \frac{f(x,y)}{f_Y(y)}$

$$F[x|y] = \Pr[X \leq x | Y = y] = \int_{-\infty}^{x} f(z|y) dz$$

- Total probability rule

$$\Pr[X \leq x] = \int_{-\infty}^{\infty} \Pr[X \leq x | Y = y] f_Y(y) dy$$

Some Common Distributions

- Uniform between $[a, b]$ $f(x) = \begin{cases} \frac{1}{b-a} & if\ a \leq x \leq b \\ 0 & otherwise \end{cases}$
- Exponential $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$
- Normal (Gaussian), $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Multi-Variable Gaussian $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}))$$

  - $\boldsymbol{x}$ and $\boldsymbol{\mu}$ are n-dimensional vectors
  - $\boldsymbol{\Sigma}$ is the $n \times n$ covariance matrix and $|\boldsymbol{\Sigma}|$ its determinant

Functions of Random Variables

- Suppose that random variables are related through $Y = g(X)$ and the cdf of X is known $F_X(x)$

$$\text{Find } F_Y(y) = \Pr[Y \le y] = \Pr[g(x) \le y]$$

- **Example**:
  - Let $Y = aX + b$, then
  - $F_Y(y) = \Pr[Y \le y] = \Pr[aX + b \le y] = \Pr\left[X \le \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$
- **Useful formula**: Let x_i be the roots of $y = g(x)$. Then

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{\left|\frac{dg}{dx}(x_i)\right|}$$

- **Example**:
  - Let $Y = X^2$, then, $x_1 = \sqrt{y}$, and $x_2 = -\sqrt{y}$, so
  - $f_Y(y) = \frac{f_X(\sqrt{y})}{|2\sqrt{y}|} + \frac{f_X(-\sqrt{y})}{|-2\sqrt{y}|} = \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right)$

Expectation / Variance

- **Continuous Random Variables**
  - Expected value $E[X] = \int_{-\infty}^{\infty} xf(x)dx$
  - Variance $\sigma^2 = E[(X - E[X])^2] =$
    $$= E[X^2 - 2XE[X] + (E[X])^2]$$
    $$= E[X^2] - (E[X])^2$$
  - Standard deviation σ.
- **Discrete Random Variables**
  - Expected value $E[X] = \sum_x x \Pr[X = x]$
- **Moments**: nth moment $E[X^n]$
- **Coefficient of Variation** $C_X = \frac{\sigma_X}{E[X]}$

Covariance and correlation

- Let X, Y be random variables with joint pdf $f(x,y)$, **covariance**

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

  - $Cov(X,Y) = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$
    $= E[XY] - E[X]E[Y]$

- **Correlation coefficient**

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- Let X be a random variable with pdf $f(x)$ then the **characteristic function** is defined as

$$\varphi_X(t) = E\left[e^{jtX}\right] = \int_{-\infty}^{\infty} e^{jtx} f(x) dx$$

Law of Large Numbers (LLN)

- Let the sequence of n i.i.d. random numbers $X_1, X_2, \dots X_n$ each with mean μ and variance σ^2, and define the sample mean

$$S_n = \frac{1}{n}\left(X_1 + X_2 + \dots + X_n\right)$$

- **Weak LLN:** Assume a small $\varepsilon > 0$, then

$$\lim_{n\to\infty} \Pr\left(\left|S_n - \mu\right| > \epsilon\right) = 0$$

- **Strong LLN**

$$\Pr\left(\lim_{n\to\infty} S_n = \mu\right) = 1$$

Central Limit Theorem (CLT)

- Let the sequence of n i.i.d. random numbers $X_1, X_2, \ldots X_n$ each with mean μ and variance σ^2, and define the sample mean

$$S_n = \frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right)$$

- Then, as n grows large, the distribution of S_n approximates the Normal distribution (Gaussian) with mean μ and variance σ^2/n.

Random Process (Stochastic Process)

- Collection of Random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ indexed by a variable t.
  - Continuous random process $\{X(t)\}$ for all $t \in \mathbb{R}$
  - Discrete time random process $\{X(t)\}$ for all $t = 0,1,2,\ldots$
- To define a random process we need the joint cdf of *all* random variables that define the process.

$$F_X(x_0, \ldots, x_n; t_0, \ldots, t_n) = \Pr[X(t_0) \leq x_0, \ldots., X(t_n) \leq x_n]$$

- Independent Process $\{X(t)\}$

$$F_X(x_0, \ldots, x_n; t_0, \ldots, t_n) = F_{X_0}(x_0; t_0) \ldots F_{X_n}(x_n; t_n)$$

- Independent Identically Distributed (iid)

$$F_X(x; t) = F_{X_0}(x_0; t_0) = \ldots = F_{X_n}(x_n; t_n)$$

Stationary Process

- **Autocorrelation:** Let the process $\{X(t)\}$ and two time instances t_1, and t_2, then the autocorrelation is given by
$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)]$$
- **Strict-sense stationary**: The process $\{X(t)\}$ exhibits the same statistical behavior at all time.
$$F_X(x_0, \ldots, x_n; t_0 + \tau, \ldots, t_n + \tau) = F_X(x_0, \ldots, x_n; t_0, \ldots, t_n)$$
for all τ.
  - $R_{XX}(t_1, t_2) = R_{XX}(t_2 - t_1)$, i.e., it does not depend on t_1, and t_2 but only on the difference $t_2 - t_1$.
  - **Ergodicity:** Ensample average is equal to time average
- **Wide–sense stationary**:
$$E[X(t)] = C \text{ (constant) for all } t.$$
$$E[X(t)X(t + \tau)] = g(\tau)$$

Gaussian (Normal) Process

- Let the process $\{X(t)\}$ be Gaussian, then the distribution of $X(t)$ and any time instant t is Gaussian

$$X(t) \sim N(\mu_t, \sigma_t^2)$$

$$f(x,t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(x-\mu_t)^2}{2\sigma_t^2}\right)$$

- The joint distribution of the points $t_1, \ldots, t_n$ is a multi-variable Gaussian $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - $\boldsymbol{x}$ and $\boldsymbol{\mu}$ are n-dimensional vectors
  - $\boldsymbol{\Sigma}$ is the $n \times n$ autocovariance matrix
- Gaussian White Noise
  - The variables $\{X(t)\}$ are independent identically distributed (i.i.d.) $X(t) \sim N(\mu, \sigma^2)$ for all t.