ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

APPLIED STATISTICS AND EXPERIMENTAL DESIGN

INTRODUCTION

Applied Statistics and Experimental Design

Assoc. Prof. Nguyễn Linh Giang

School of Information and Communication Technology

giangnl@soict.hust.edu.vn

II. Basics of Probability theory

- 2.1. Probability.
- 2.2. Random variables.
- 2.3. Law of large number
- 2.4. Normal probability distribution (Gaussian distribution).

2.1. Probability

- Probability
  - Laplace classic definition of probability:

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}},$$

  - Relative frequency definition of probability:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

2.1. Probability

- Kolmogorov axiomatic formulation
  - Ω: sample space: set of all experimental outcomes

    $$\Omega = \{ \xi_1, \xi_2, \dots \xi_k, \dots, \xi_n, \dots \}$$

  - Event – any subset of Ω. Number of subset of sample space : 2^n if $n < \infty$.
  - σ-field *F* of subsets of Ω
  - P: a probability measure on the sets in *F*
    - A – any event
    - 3 axiom of probability

$$\text{(i)} \quad P(A) \geq 0 \quad \text{(Probability is a nonnegative number)}$$

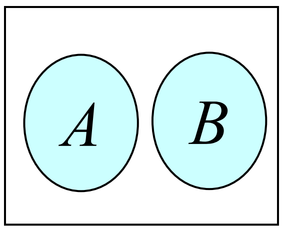
$$\text{(ii)} \quad P(\Omega) = 1 \quad \text{(Probability of the whole set is unity)}$$

$$\text{(iii)} \quad \text{If } A \cap B = \phi, \quad \text{then} \quad P(A \cup B) = P(A) + P(B).$$

  - $< \Omega, F, P >$: probability model

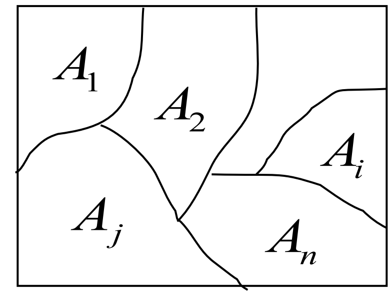
2.1. Probability

- Events: A and B
  - Mutualy exclusive events: $A \cap B = \varnothing$
  - Partition of Ω:



$A \cap B = \phi$

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1}^{n} A_i = \Omega$$



  - Example: experiment of tossing two coins simultaneously
    - Elementary events:

$$\xi_1 = (H,H), \quad \xi_2 = (H,T), \quad \xi_3 = (T,H), \quad \xi_4 = (T,T)$$

    - The subset $A = \{ \xi_1, \xi_2, \xi_3 \}$

2.1. Probability

- Conditional probability and independence
  - N independent trials,
  - N_A, N_B, N_{AB} : the number of times events A, B and AB occur.
  - For large N $\quad P(A) \approx \dfrac{N_A}{N}, \quad P(B) \approx \dfrac{N_B}{N}, \quad P(AB) \approx \dfrac{N_{AB}}{N}.$
  - Conditional probability: P(A|B)

$$P(A \mid B) = \frac{N_{AB}}{N_B} = \frac{N_{AB} / N}{N_B / N} = \frac{P(AB)}{P(B)}$$

2.1. Probability

- Properties of conditional probability
  - P(A|B) is nonnegative:

$$P(A \mid B) = \frac{P(AB) \geq 0}{P(B) > 0} \geq 0,$$

  - P(Ω|B) = 1

$$P(\Omega \mid B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1,$$

  - If A∩C = ∅ ,

$$P(A \cup C \mid B) = P(A \mid B) + P(C \mid B),$$

2.1. Probability

- If B $\subset$ A then P(A|B) = 1
- If A $\subset$ B then P(A|B) > P(A)
- Let, A_1, A_2, …, A_n are pair wise disjoint and their union is Ω:

$$A_i \cap A_j = \varnothing, \qquad \bigcup_{i=1}^{n} A_i = \Omega.$$

  - B is an event

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

2.1. Probability

- Independence: *A* and *B* events

  P(AB) = P(A) P(B)

  - If A and B are independent events:

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

- Bayes theorem

$$P(A \mid B) = \frac{P(B \mid A)}{P(B)} \cdot P(A)$$

- Generalized Bayes theorem:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)},$$

2.1. Probability

- Bayes' theorem interpretation:
  - *P*(*A*) represents the a-priori probability of the event *A*.
  - The event *B* is new knowledge obtained from an experiment.
  - Conditional probability *P(A|B)* of A given B –
  a-posteriori probability
  - The new information should be used to improve knowledge of *A*.

2.1. Probability

- Example:
  - In a box: 6 white and 4 black balls.
  - Remove two balls randomly without replacement.
  - P{the first one is white and the second one is black} = ?
  - Question: are events W_1 and B_2 independent ?
  - Solution:
    - W1 = “first ball removed is white”
    - *B*2 = “second ball removed is black”
    - $P(W_1 \cap B_2)$ = ?

2.1. Probability

- Example: Two boxes $B1$ and $B2$ contain 100 and 200 light bulbs respectively. The first box ($B1$) has 15 defective bulbs and the second - 5. Suppose a box is selected at random and one bulb is picked out.
  - (a) What is the probability that it is defective?
  - (b) Suppose we test the bulb and it is found to be defective. What is the probability that it came from box 1?

2.1. Probability

- Repeated trials, Bernoulli trials
  - Consider n independent experiments with models (Ω_1, F_1, P_1), (Ω_2, F_2, P_2), ..., (Ω_n, F_n, P_n).
    - Let $\xi_1 \in \Omega_1$, $\xi_2 \in \Omega_2$,..., $\xi_n \in \Omega_2$: elementary events.
    - A joint performance of the n experiments produces an elementary events $\omega = (\xi_1, \xi_2, ..., \xi_n)$.
    - Consider space $\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_n : \xi_1 \in \Omega_1, ..., \xi_n \in \Omega_n,$
    - Events in combined space Ω are of the form $A_1 \times A_2 \times ... \times A_n$.
    - If n experiments are independent, then $P(A_1 \times A_2 \times ... \times A_n) = P(A_1) \times ... \times P(A_n)$

2.1. Probability

- An event *A* has probability *p* of occurring in a single trial. Find the probability that *A* occurs exactly *k* times in determined location, $k \le n$ in *n* trials.

$$P_0(\omega) = P(\{\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_n}\}) =$$
$$= P(\{\xi_{i_1}\})P(\{\xi_{i_2}\})\cdots P(\{\xi_{i_k}\})\cdots P(\{\xi_{i_n}\}) =$$
$$= \underbrace{P(A)P(A)\cdots P(A)}_{k}\ \underbrace{P(\overline{A})P(\overline{A})\cdots P(\overline{A})}_{n-k} = p^k q^{n-k}.$$

- P{ A occurs exactly *k* time in *n* trials } = $C^k_n p^k q^{n-k}$
  - Bernoulli formula.

2.1. Probability

- De Moivre – Laplace Theorem
  - Guess that: $n \to \infty$ with fixed p.
  - *k* is in the neighborhood $\sqrt{npq}$ of *np*.
  - Bernoulli probability estimation:

$$C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}.$$

- Stirling formula for n! approximation:

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$$

2.1. Probability

- Estimation of Bernoulli formula

$$\binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!k!} p^k q^{n-k},$$

$$\binom{n}{k} p^k q^{n-k} > c_1 \sqrt{\frac{n}{2\pi(n-k)k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

$$\binom{n}{k} p^k q^{n-k} < c_2 \sqrt{\frac{n}{2\pi(n-k)k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

- Các hằng số c_1 và c_2 khá gần nhau.

$$c_1 = e^{\left\{\frac{1}{12n+1} - \frac{1}{12(n-k)} - \frac{1}{12k}\right\}} \qquad c_2 = e^{\left\{\frac{1}{12n} - \frac{1}{12(n-k)+1} - \frac{1}{12k+1}\right\}}.$$

2.1. Probability

- Example:
  - Toss a coin *n* times. Probability of getting *k* heads in *n* trials = ?
  - Consider rolling a fair dice eight times. Find the probability that either 3 or 4 shows up five times.
  - Suppose 5,000 components are ordered. The probability that a part is defective equals 0.1. What is the probability that the total number of defective parts does not exceed 400 ?

2.1. Probability

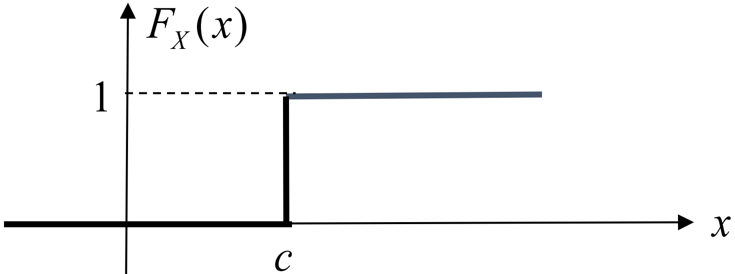
- Homeworks
  - For each of the following, explain why the events A and B are or are not independent.
    - Consider the population of HUST undergraduate students from which one student is selected at random. Let A denote the event that the student is female and let B denote the event that the student is concentrating in education.
    - Consider the population of registered voters, from which one voter is selected at random. Let A denote the event that the voter belongs to a country club and let B denote the event that the voter is a Republican

2.2. Random variables

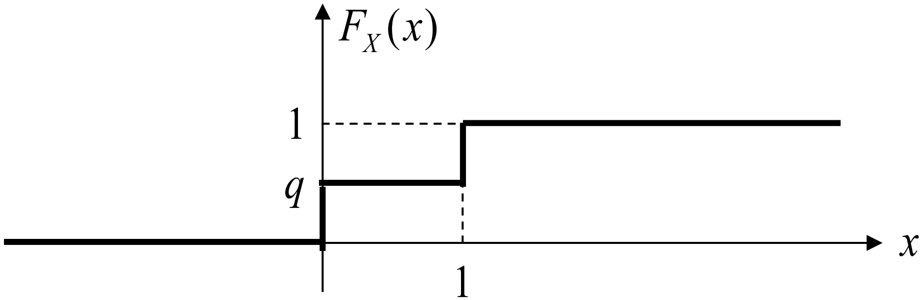
- Definition
  - (Ω, F, P) - probability model for an experiment,
  - X - a function that maps every $\xi \in \Omega$ to a unique point $x \in R$
  - **Random Variable (r.v)**: a finite single valued function that maps the set of all experimental outcomes Ω into the set of real numbers R is said to be a r.v, if the set $A=\{\xi | X(\xi) \leq x\}$ is an event $A \subset F$ for every x in R.

2.2. Random variables

- Probability distribution function
  - Probability distribution function of random variable X: (pdf)
  - $F_X(x)=P\{\xi|X(\xi)\leq x\} \geq 0$
  - Examples
    - X – r.v: $X(\xi) = c$, $\xi\in\Omega$, $F_X(x) = ?$



    - X – r.v: coin tossing, $\Omega = \{H, T\}$. X(T)=0, X(H)=1. $F_X(x) = ?$



2.2. Random variables

- Properties of probability distribution function
  - $F_X(x)$ is a distribution function:
    - $F_X(-\infty) = 0$; $F_X(+\infty) = 1$
    - Pdf is nondecreasing function:

      if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$
    - Pdf is right continuous: $F_X(x^+) = F_X(x)\ \forall x$

2.2. Random variables

- If $F_X(x_0) = 0$ for some x_0,

  then $F_X(x) = 0 \; \forall x \leq x_0$

- $P\{ X(\xi) > x \} = 1 - F_X(x)$

- If $x_2 > x_1$, then

  $P\{ x_1 < X(\xi) \leq x_2 \} = F_X(x_2) - F_X(x_1)$

- $P\{ X(\xi) = x \} = F_X(x) - F_X(x^-)$

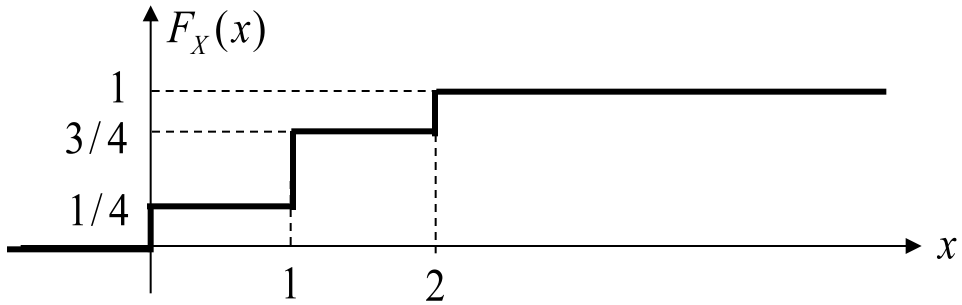
2.2. Random variables

- Continuous and discrete random variables
  - X – continuous r.v if $F_X(x)$ is continuous
    - For continuous r.v, $F_X(x^-) = F_X(x)$ and $P\{X=x\} = 0$
  - If $F_X(x)$ = const, except for a finite number of jump discontinuities, then *X* is said to be a discrete-type r.v.
    - If x_i is such a discontinuity point, then

$$p_i = P\{ X = x_i \} = F_X(x) - F_X(x^-)$$

2.2. Random variables

- Examples
  - A fair coin is tossed twice, and let the r.v *X* represent the number of heads. Find $F_X(x)$



2.2. Random variables

- Distribution density function
  - Derivative of distribution function $F_X(x)$ is density function $f_X(x)$ of the r.v.

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

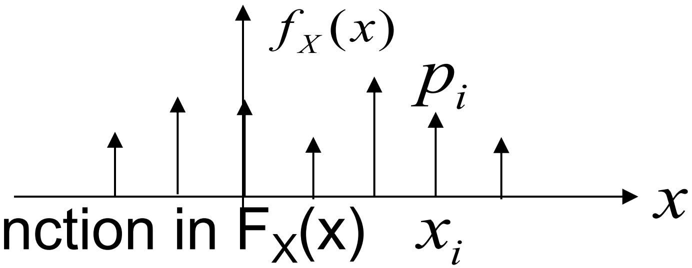
  - $F_X(x)$ is monoton nondecreasing function, so that

$$f_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \geq 0,$$

2.2. Random variables

- If r.v is continuous then $f_X(x)$ is continuous.
- For discrete r.v

$$f_X(x) = \sum_i p_i \delta(x - x_i),$$



- x_i is jump-discontinuity point function in $F_X(x)$
-

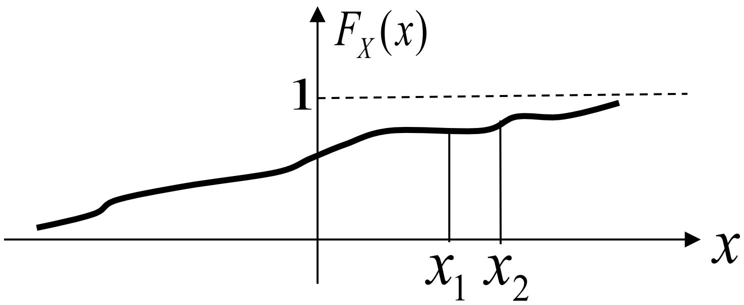
$$F_X(x) = \int_{-\infty}^{x} f_x(u)du.$$

$$F_X(+\infty) = 1, \qquad \int_{-\infty}^{+\infty} f_x(x)dx = 1,$$

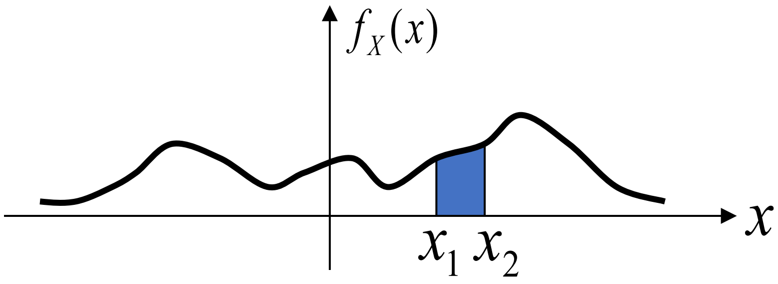
$$P\{ x_1 < X(\xi) \le x_2 \} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx.$$

2.2. Random variables

$$P\{\,x_1 < X(\xi) \le x_2\,\} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx.$$



(a)

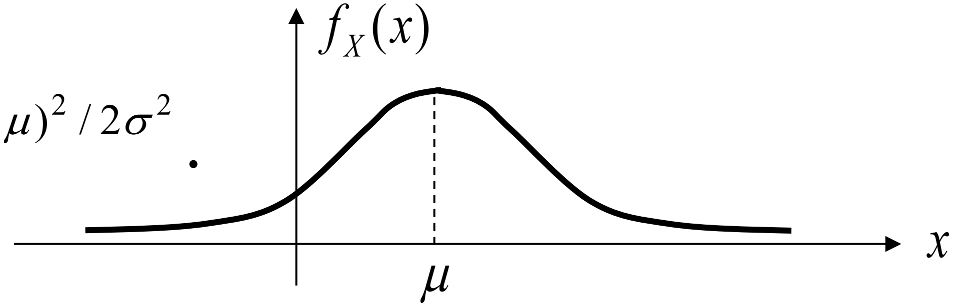


(b)

2.2. Random variables

- Some continuous random variables
  - Normal (Gaussian) random variables $X \sim N(\mu, \sigma^2)$
    - Density function

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



    - Bell shape curve
    - Distribution function

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

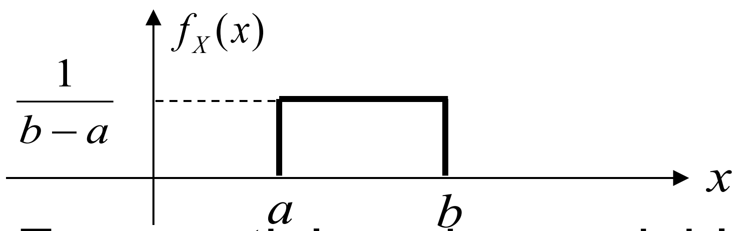
$$G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

2.2. Random variables

- Uniform random variables:
  - Density function

$$X \sim U(a,b),\ a < b,$$

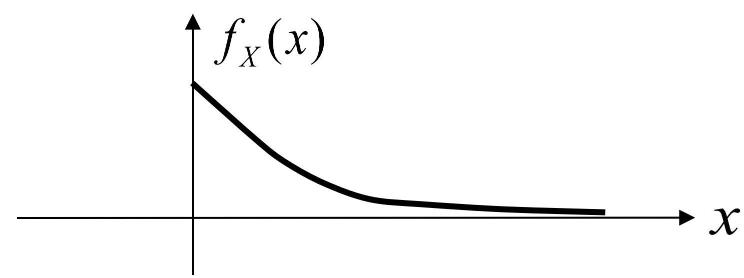
$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$



- Exponential random variables
  - Density function

$$X \sim \varepsilon(\lambda)$$

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



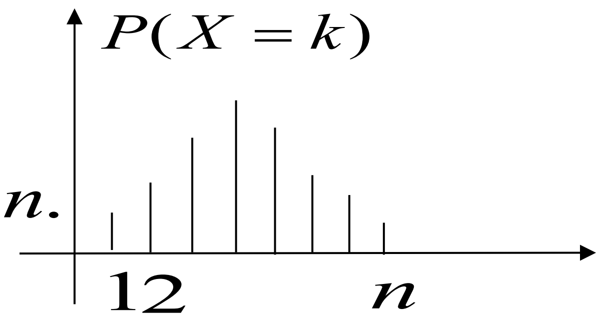
2.2. Random variables

- Some important discrete random variables
  - Bernoulli r.v: X takes values 0, 1

$$P(X=0)=q, \quad P(X=1)=p.$$

  - Binomial r.v $\qquad X \sim B(n,p),$

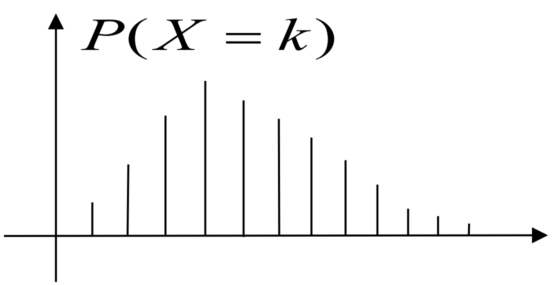
$$P(X=k)=\binom{n}{k}p^k q^{n-k},\ k=0,1,\cdots,n.$$



  - Poisson r.v

$$X \sim P(\lambda),$$

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\ k=0,1,\cdots,\infty.$$



2.3 Some characteristics of random variables

- Mean value

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx.$$

  - For discrete random variable

$$\eta_X = \overline{X} = E(X) = \int x \sum_i p_i \delta(x - x_i)dx = \sum_i x_i p_i \underbrace{\int \delta(x - x_i)dx}_{1}$$

$$= \sum_i x_i p_i = \sum_i x_i\, P(X = x_i)\,.$$

  - Example: uniform random variable
  - Example: exponential random variable

2.3 Some characteristics of random variables

- Variance
  - For r.v X with mean μ
  - Variance

$$\sigma_X^2 = E[(X-\mu)^2] > 0.$$

  - Or

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx > 0.$$

  - Standard deviation

$$\sigma_X = \sqrt{E(X-\mu)^2}$$

  - Variance anf means

$$Var(X) = \sigma_X^2 = E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2] =$$

$$E[X^2] - E[2X\mu] + E[\mu^2] = E[X^2] - 2E[X]\mu + \mu^2$$

$$= E(X^2) - [E(X)]^2 = \overline{X^2} - \overline{X}^2.$$

2.3 Some characteristics of random variables

- Example: variance of Poisson r.v.
- Example: variance of Gaussian r.v.
- Example

2.3 Some characteristics of random variables

- Moments
  - X – random variables

$$m_n = \overline{X^n} = E(X^n), \quad n \geq 1$$

  - Central moments

$$\mu_n = E[(X-\mu)^n]$$

  - Relation between moments and central moments

$$\mu_n = E[(X-\mu)^n] = \sum_{k=0}^{n} C_n^k \ m_k \ (-\mu)^{n-k}.$$

  - Mean and variance:

$$\mu = m_1, \quad \sigma^2 = \mu_2.$$

2.3 Some characteristics of random variables

- Generalized moments of X about *a*

$$E[(X-a)^n]$$

- Absolute moments of X

$$E[|X|^n]$$

2.3 Some characteristics of random variables

- Characteristic function
  - Continuous r.v X

$$\Phi_X(\omega) \stackrel{\Delta}{=} E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)dx.$$

    - We have:

$$\Phi_X(0) = 1,$$

    - And

$$\left|\Phi_X(\omega)\right| \le 1 \qquad \forall \omega$$

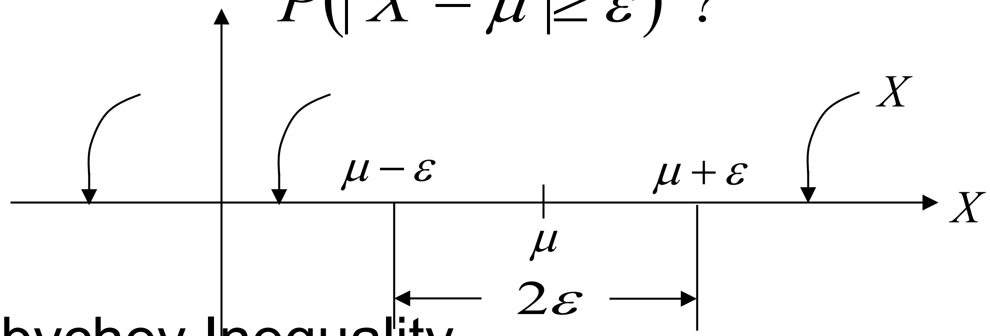
  - Discrete r.v X

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k).$$

2.4 Chebychev Inequality and Law of Large numbers

- Chebychev Inequality
  - Consider an interval of width 2ε symmetrically centered around its mean μ

$$P(|X-\mu| \geq \varepsilon)\ ?$$



  - Chebychev Inequality

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2},$$

2.4 Chebychev Inequality and Law of Large numbers

- Weak law and strong law of large numbers
  - X_i - independent, identically distributed Bernoulli random variables:

$$P(X_i) = p, \qquad P(X_i = 0) = 1 - p = q,$$

  - k = $X_1 + X_2 + ... + X_n$ – number of successes in *n* trials
  - Weak law of large numbers:

$$P\left\{ \left| \frac{k}{n} - p \right| > \varepsilon \right\} \leq \frac{pq}{n\varepsilon^2}.$$

  - Strong law of large numbers:
    - The ratio *k/n* tends to *p* not only *in probability*, but *with probability* 1

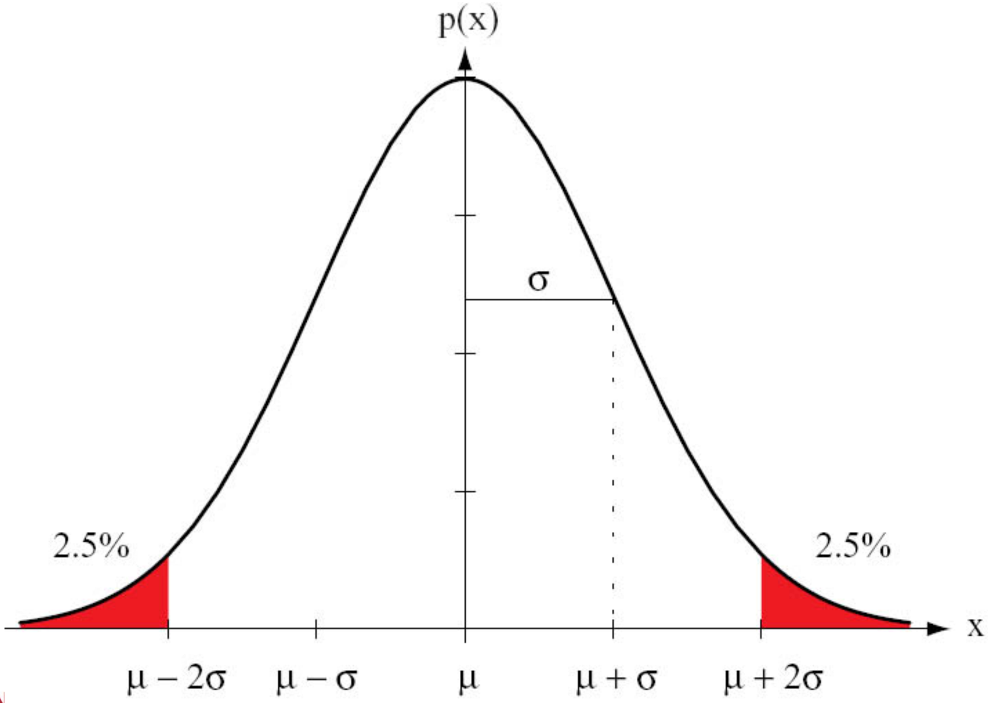
2.5. Gaussian distribution

- Gaussian distribution
  - One dimentional normal distribution

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

$$\mu \equiv E(x) = \int_{-\infty}^{\infty} x p(x) dx$$

$$\sigma^2 \equiv E((x-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx$$



2.5. Gaussian distribution

- Multidimentional Gaussian distribution (*d* dimentional)

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\left|\sum\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T \sum{}^{-1} (\mathbf{x}-\mu)\right]$$

$$\mu \equiv E(\mathbf{x}) = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x}$$

- Where, **x**, **μ** are *d*-dimentional vector
- ∑ covariance matrix - symmetric and semi-positive defined matrix.
- Consider cases, when ∑ positive defined.

$$\Sigma = E((\mathbf{x}-\mu)(\mathbf{x}-\mu)^T) = \int (\mathbf{x}-\mu)(\mathbf{x}-\mu)^T p(\mathbf{x}) d\mathbf{x}$$

$$\mu_i = E(x_i)$$

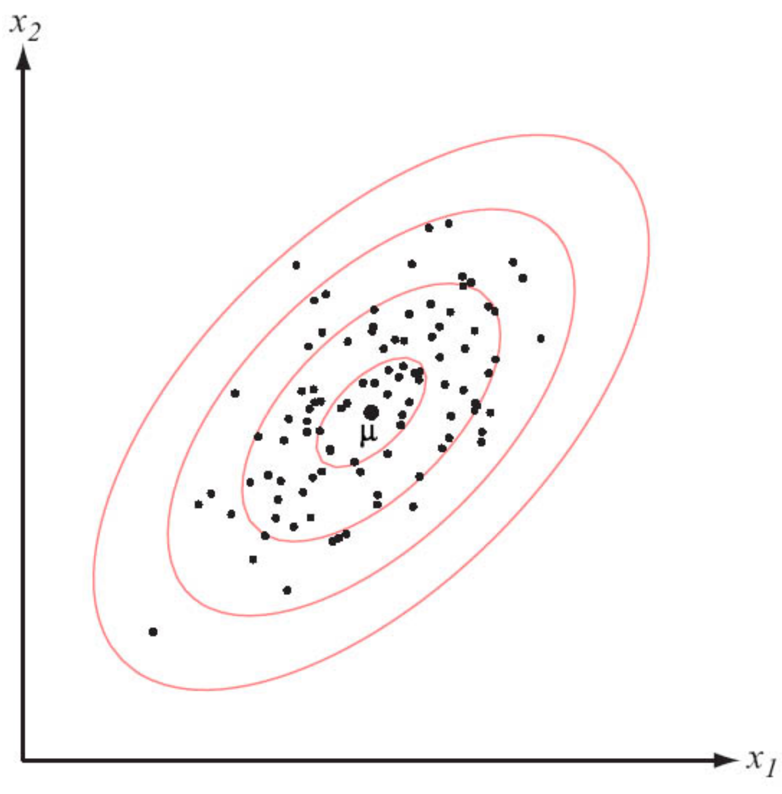
$$\sigma_{ij} = E((x_i - \mu_i)(x_j - \mu_j))$$

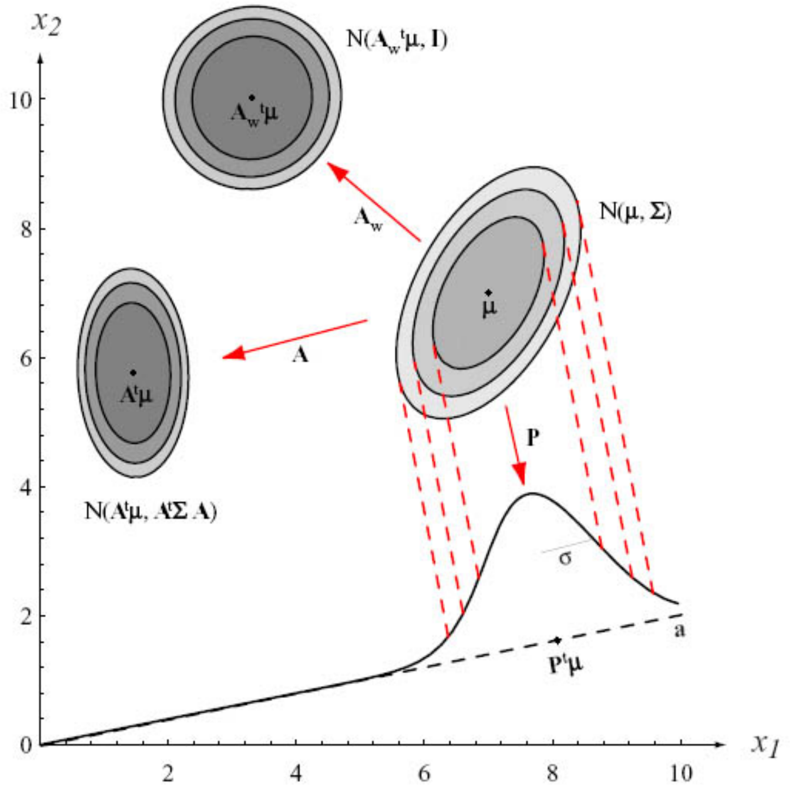
2.5. Gaussian distribution

- Covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{00} & & & \\ & \sigma_{11} & \sigma_{ij, i \neq j} & \\ & & \sigma_{22} & \\ & & & \sigma_{33} \end{bmatrix}$$

2.5. Gaussian distribution





2.5. Gaussian distribution

- Multidimentional Gaussian distribution is fully defined by d+d(d+1)/2 parameters of mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.
- Mahalanobis distance:

$$r = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

ĐẠI HỌC BÁCH KHOA

25 YEARS ANNIVERSARY SOICT

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for your attentions!

soict.hust.edu.vn/ fb.com/groups/soict