PASFML AI45 Mansoura, Session 3 23/12/2024

1- Covariance matrix (statistical tools)

2- Discrete Probability distributions (PMF, CDF

- Statistical tools for distrete distributions.

- Intro. to continuous distributions.

---

1 Review

mean: $\bar{X} = \mu_X = \sum_i (x_i \; P_X(x_i))$

Variance: $\sigma_X^2 = \overline{(x_i - \bar{X})^2} = \sum_i (x_i - x)^2 \, P_X(x_i)$

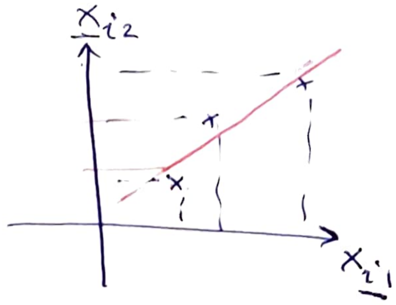
$= \overline{X^2} - \bar{X}^2 \equiv \overline{X^2} - \mu_X^2$

Covariance $\sigma_{xy} = \overline{(x_i - \bar{X})(y_i - \bar{Y})}$

Covariance Matrix "C"

Correlation Matrix "R"

$\vec{X}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}$ $C = \begin{bmatrix} \sigma_{x_{i1}}^2 & \sigma_{x_{i1} x_{i2}} \\ \sigma_{x_{i1} x_{i2}} & \sigma_{x_{i2}}^2 \end{bmatrix}$



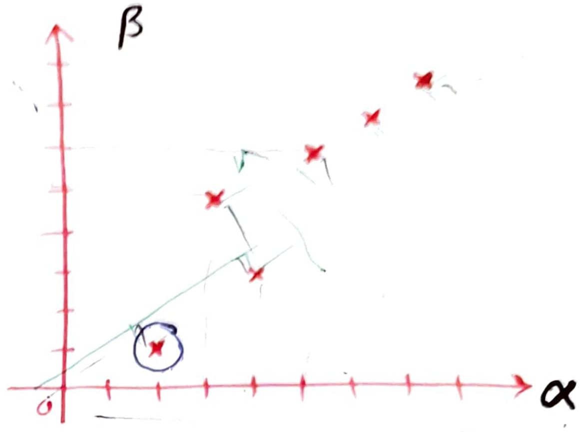
$\vec{X}_i = \begin{array}{cc} \begin{bmatrix} \cancel{x}_{i1} \\ \cancel{x}_{i2} \\ \vdots \\ \cancel{x}_{im} \end{bmatrix} & \begin{array}{c} X \\ Y \\ \\ Z \end{array} \end{array}$

$C = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}$

$R = \begin{bmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{XY} & 1 & \rho_{YZ} \\ \rho_{XZ} & \rho_{YZ} & 1 \end{bmatrix}$

$\vec{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $\vec{x}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ $\vec{x}_3 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ $\vec{x}_4 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$ $\vec{x}_5 = \begin{bmatrix} 6 \\ 7 \end{bmatrix}$ $\vec{x}_6 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$

$$X = \begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 5 & 3 & 6 & 7 & 8 \end{bmatrix}$$



$$\boxed{X = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 4 & 3 \\ 5 & 6 \\ 6 & 7 \\ 7 & 8 \end{bmatrix}}$$

1. $$\mu_\alpha = \frac{2+3+4+5+6+7}{6} = 4.5$$

$$\mu_\beta = \frac{1+5+3+6+7+8}{6} = 5$$

$$X_c = \begin{bmatrix} \alpha_1 - \mu_\alpha & \alpha_2 - \mu_\alpha & \cdots \\ \beta_1 - \mu_\beta & \cdots & \end{bmatrix} \Leftarrow \sigma_\alpha^2 = \frac{(2-4.5)^2 + \cdots + (7-4.5)^2}{(n) \text{ or } (n-1)} = \checkmark$$

$(n) \to 6$ $\quad (n-1) \to 5$

to be discussed in statistics session.

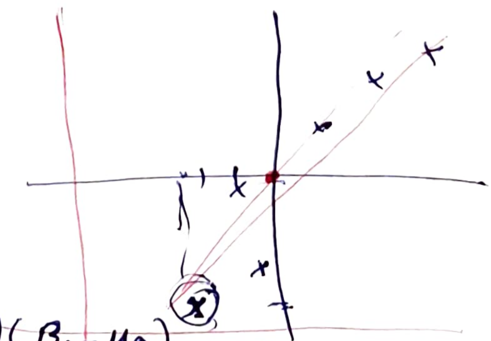
centered X

$$\sigma_\beta^2 = \checkmark$$

$$X_c = \begin{bmatrix} 2-4.5 & 3-4.5 & 4-4.5 & 5-4.5 & 6-4.5 & 7-4.5 \\ 1-5 & 5-5 & 3-5 & 6-5 & 7-5 & 8-5 \end{bmatrix}$$

$$X_c = \begin{bmatrix} \alpha_i - \mu_\alpha \\ \beta_i - \mu_\beta \end{bmatrix}$$

$[\,a \quad b$



$$\frac{1}{n} X_c X_c^T = \frac{1}{n} \begin{bmatrix} \sum_i (\alpha_i - \mu_\alpha)^2 & \sum (\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta) \\ \sum_i (\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta) & \sum_i (\beta_i - \mu_\beta)^2 \end{bmatrix} = C$$

or $(n-1)$

$$X_c = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} \bigcirc & \\ & \end{bmatrix}$$

$$\begin{bmatrix} \alpha_1 - \mu_\alpha & \alpha_2 - \mu_\alpha & \alpha_3 - \mu_\alpha \\ \beta_1 - \mu_\beta & \beta_2 - \mu_\beta & \beta_3 - \mu_\beta \end{bmatrix} \begin{bmatrix} \alpha_1 - \mu_\alpha & \beta_1 - \mu_\beta \\ \alpha_2 - \mu_\alpha & \beta_2 - \mu_\beta \\ \alpha_3 - \mu_\alpha & \beta_3 - \mu_\beta \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i (\alpha_i - \mu_\alpha)(\alpha_i - \mu_\alpha) & \sum_i (\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta) \\ \sum_i (\beta_i - \mu_\beta)(\alpha_i - \mu_\alpha) & \sum_i (\beta_i - \mu_\beta)(\beta_i - \mu_\beta) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i (\alpha_i - \mu_i)^2 & \sum_i (\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta) \\ \sum_i (\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta) & \sum_i (\beta_i - \mu_\beta)^2 \end{bmatrix}$$

$$\boxed{C = \frac{1}{n} X_c X_c^T}$$

(or $\frac{1}{n-1}$, sample variance, bias correction)

→ if data are in row vectors of data matrix.

$$C = \frac{1}{n} X_c^T X_c$$

$$\underline{C} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}$$

Covariance matrix

$$\underline{R} = \begin{bmatrix} 1 & \rho_{\alpha\beta} \\ \rho_{\alpha\beta} & 1 \end{bmatrix}$$

Correlation Matrix

$$X_n = \begin{bmatrix} \frac{\alpha_i - \mu_\alpha}{\sigma_\alpha} \\ \frac{\beta_i - \mu_\beta}{\sigma_\beta} \end{bmatrix} \qquad X_n^T \quad \begin{bmatrix} \quad \end{bmatrix}$$

$$\frac{1}{n} X_n X_n^T = \cancel{\frac{1}{n}} \begin{bmatrix} \boxed{\sum_i \frac{(\alpha_i - \mu_\alpha)^2}{n\,\sigma_\alpha^2}} & \boxed{\frac{1}{n} \sum_i \frac{(\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta)}{\sigma_\alpha \sigma_\beta}} \\ \boxed{\quad} & \frac{1}{n} \frac{\sum (\beta_i - \mu_\beta)^2}{\sigma_\beta^2} \end{bmatrix}$$

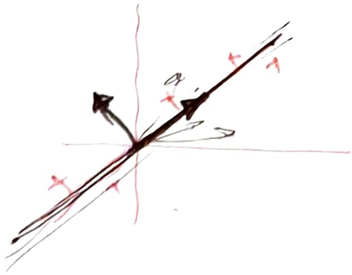
$$= \begin{bmatrix} 1 & \rho_{xy} \\ \rho_{xy} & 1 \end{bmatrix} = R$$

e.g.,

$$\frac{1}{n} \begin{bmatrix} \frac{x_1 - \mu_x}{\sigma_x} & \frac{x_2 - \mu_x}{\sigma_x} & \frac{x_3 - \mu_x}{\sigma_x} \\ \frac{y_1 - \mu_y}{\sigma_y} & \frac{y_2 - \mu_y}{\sigma_y} & \frac{y_3 - \mu_y}{\sigma_y} \end{bmatrix} \begin{bmatrix} \frac{x_1 - \mu_x}{\sigma_x} & \frac{y_1 - \mu_y}{\sigma_y} \\ \frac{x_2 - \mu_x}{\sigma_x} & \frac{y_2 - \mu_y}{\sigma_y} \\ \frac{x_3 - \mu_x}{\sigma_x} & \frac{y_3 - \mu_y}{\sigma_y} \end{bmatrix}$$

$$= \begin{bmatrix} \overbrace{\boxed{\frac{1}{n} \sum_i \frac{(x_i - \mu_x)^2}{\sigma_x^2}}}^{\sigma_x^2} & \boxed{\frac{1}{n} \sum_i \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y}} \sim \sigma_{xy} \\ \boxed{\quad} & \boxed{\frac{1}{n} \sum_i \frac{(y_i - \mu_y)^2}{\sigma_y^2}} \leftarrow \sigma_y^2 \end{bmatrix}$$

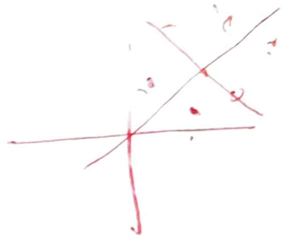
$$X \Rightarrow X_c \Rightarrow C = \begin{bmatrix} \sigma_\alpha & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta \end{bmatrix}$$



eigenvectors of the Covariance matrix are Principle components of data.

$$\vec{v}_1 \;,\; \vec{v}_2$$

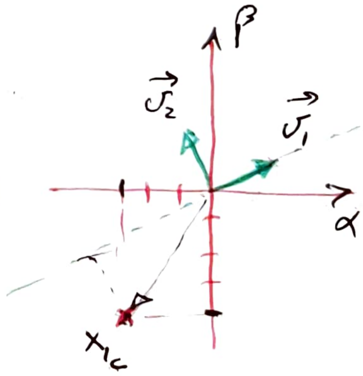
$$\lambda_1 \;,\; \lambda_2$$



$$C = \begin{bmatrix} 3.5 & 4.25 \\ 4.5 & 8 \end{bmatrix}$$

$$\vec{v}_1 \approx \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix} \qquad \vec{v}_2 \approx \begin{bmatrix} -0.82 \\ 0.57 \end{bmatrix}$$

$$\lambda_1 \approx 11.36 \qquad \lambda_2 \simeq 0.14$$



$$\vec{X}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \vec{X}_{1_c} = \begin{bmatrix} 2-\mu_\alpha \\ 1-\mu_\beta \end{bmatrix} = \begin{bmatrix} 2-4.5 \\ 1-5 \end{bmatrix} = \begin{bmatrix} -2.5 \\ -4 \end{bmatrix}$$

$$\left\| \mathrm{Proj}_{\vec{v}_1}(\vec{X}_{1_c}) \right\| = \left\| \frac{\vec{X}_{1_c} \cdot \vec{v}_1}{\underbrace{\vec{v}_1 \cdot \vec{v}_1}_{1}} \cancel{\vec{v}_1} \right\| = \begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix} = -4.7$$

$$\boxed{\vec{X}_{1_c} \approx -4.7\, \vec{v}_1} + \cancel{\quad \vec{v}_2}$$

$$\vec{X}_{1_c} \approx -4.7 \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix} = \begin{bmatrix} -2.67 \\ -3.85 \end{bmatrix} \Rightarrow \vec{X}_1 \approx \begin{bmatrix} \_\_\_ + \mu_\alpha \\ \_\_\_ + \mu_\beta \end{bmatrix}$$

PCA ; eigendecomposition of covariance matrix $\underline{C}$

$$\underline{C} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

$$\begin{bmatrix} \sigma_x^2 & & & \\ \sigma_{xy} & \sigma_y^2 & & \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 & \\ \sigma_{xw} & & & \sigma_w \end{bmatrix}$$

$$\underline{C} = \frac{1}{n} \underline{X}\,\underline{X}^T$$

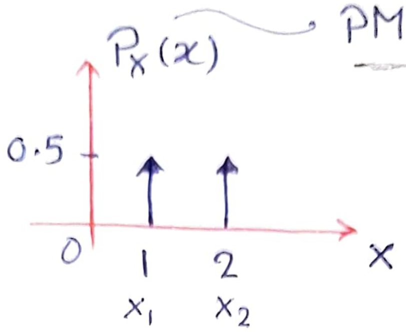
PCA ; SVD of $\underline{X}$

SVD of matrix A ; eigendecomp. of $A\underline{A}^T$

Discrete Probability Distributions

→ Random variable X

$P_X(x)$



PMF, Prob.

→ Uniform ✓

→ Bernoulli ✓

→ Binomial ✓

→ Poisson ✓

→ Geometric

→ hypergeometric

→ ....

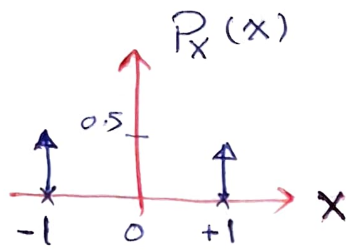
→ mean, variance of distributions

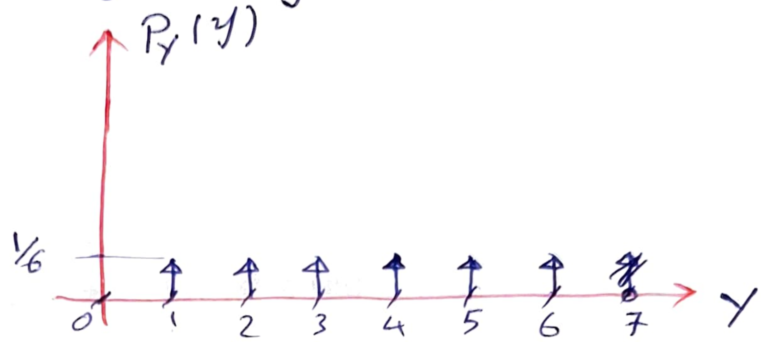
→ CDF ; cummulative distribution function.

transition towards discussing continuous distributions.

→ uniform (discrete) distribution.

→ all ~~values~~ outcomes are equally likely "equiprobable"

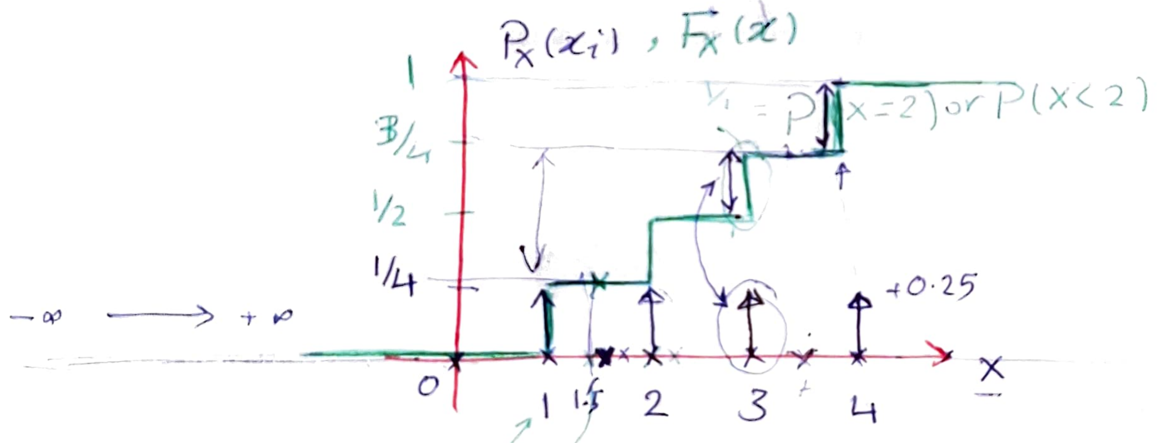


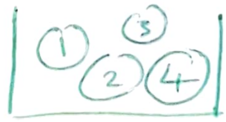


if Z is uniformly distributed R.V.

$$P_Z(z) = \begin{cases} 1/n & \text{for } i = 1:n \\ 0 & \text{otherwise} \end{cases}$$

→ CDF $F_X(x_i) = P(X \leq x) = P_X(x_i) + P_X(X < x_i)$





$n = 4$

$P_X(x_i) = \frac{1}{n} = \frac{1}{4}$

$i = 1:4$

$\sum_{i=1}^{n} \frac{1}{n} = 1$

$P_X(1.3) = 0$

$F_X(X \leq x_i) = P(X \leq 1.3) = P_X(1.3) + P(X < 1.3)$

zero   1/4

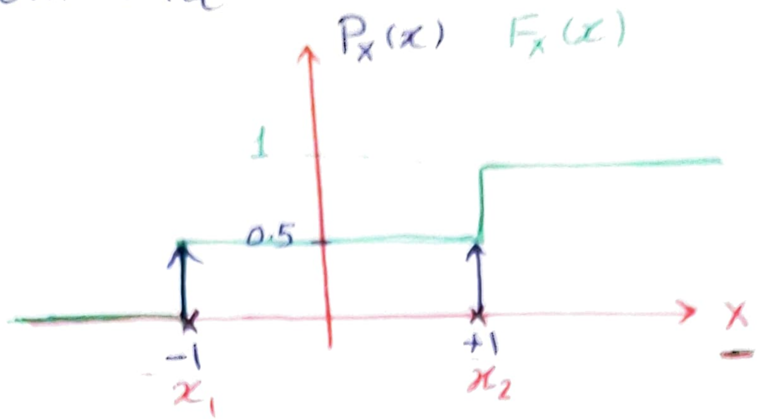
→ $F_X(-\infty) = 0$ → $F_X(+\infty) = +1$

→ nondecreasing ; if $x_2 > x_1 \Rightarrow F_X(x_2) \geq F_X(x_1)$

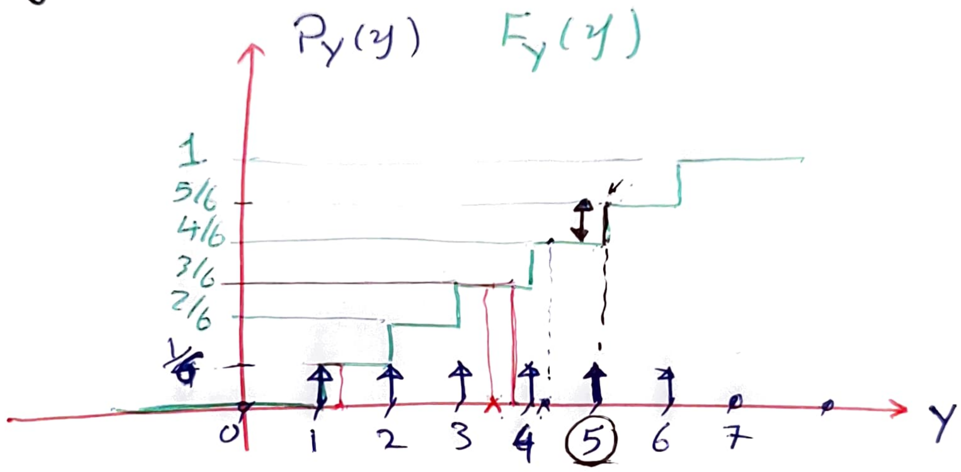
→ $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

examples

→ tossing a coin once.



→ rolling a die once,



$P_X(5) = F_X(5^+) - F_X(5^-) = 5/6 - 4/6 = 1/6$

$P_X(4.2) = F_X(4.2^+) - F_X(4.2^-) = 4/6 - 4/6 = 0$

$P(X \leq 3.5) = F_X(3.5) = 3/6$

1, 2, 3

$P(1.1 < X \leq 3.9) = F_X(3.9) - F_X(1.1) = 3/6 - 1/6 = 2/6$

1	2	3	4	5	6

2/6

if X is (discrete) uniformly distributed R.V.

$$P_X(x_i) = \begin{cases} \frac{1}{n} & \text{for } i = 1:n \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_x \equiv \bar{X} = \sum_{i=1}^{n} x_i \, P_i(x) = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma_X^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \, P_X(x_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

---

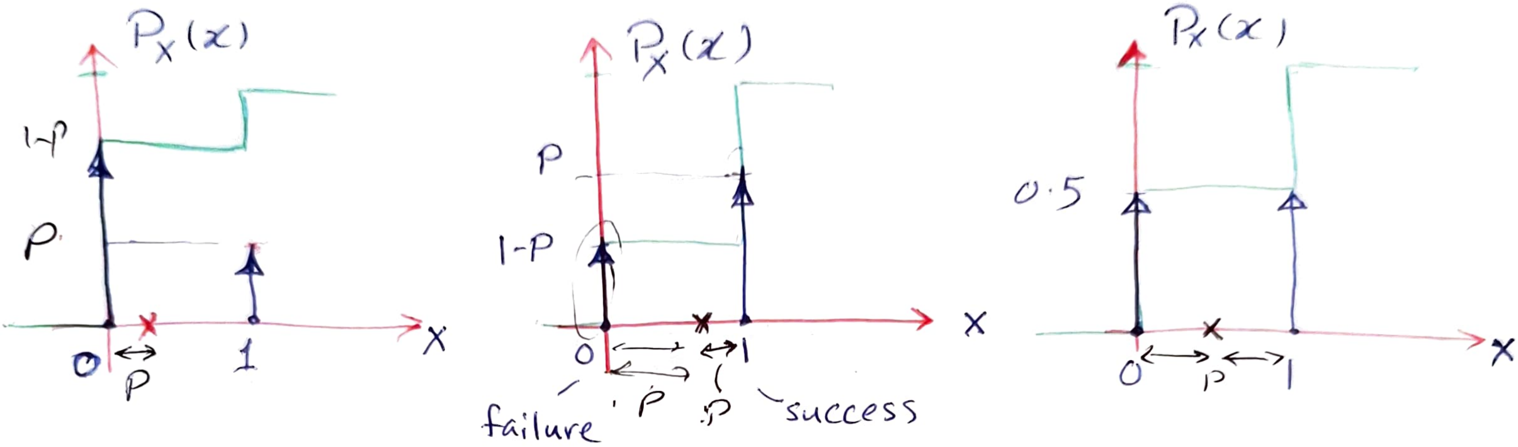
ex. rolling a die once

$$\bar{X} = 3.5 = \frac{1+2+3+4+5+6}{n}$$

$$\sigma_X^2 = \frac{1}{6} \sum_{i=1}^{6} (x_i - 3.5)^2 = \_\_\_$$

Bernoulli distribution

$$P_X(x_i) = \begin{cases} p & \text{if } x_i = 1 \\ 1-p\ (q) & x_i = 0 \\ 0 & \text{otherwise} \end{cases}$$



$$\bar{X} = \sum_i x_i P_X(x_i)$$

$\bar{X}$: weigthed average

$$= 0 \times P_X(0) + 1 \times P_X(1)$$

$$= 0 \times (1-p) + 1 \times p = p = 1-q$$

$$\sigma_X^2 = \sum_i (x_i - \bar{X})^2 P_X(x_i)$$

$$= (0-p)^2 \times (1-p) + (1-p)^2 \times p$$

$$= \_\_\_\_\_\_\_\_ \quad ? \quad p \quad q$$

$$\boxed{\begin{array}{l} p = 1-q \\ q = 1-p \\ q+p = 1 \end{array}}$$

$$\sigma_X^2 = \overline{X^2} - \bar{X}^2 \qquad p^2 = (1-q)^2$$

$$\overline{X^2} = \sum_i (x_i)^2 P_X(x) = ?$$

- Binomial distribution

<u>models</u>: distribution of # number of successes in "n" Bernoulli trials.

e.g., MCQ

1 → Ⓐ Ⓑ Ⓒ Ⓓ ←
2
⋮
10

probability of answering 1 question correctly = 1/4

P(K successes in n trials)

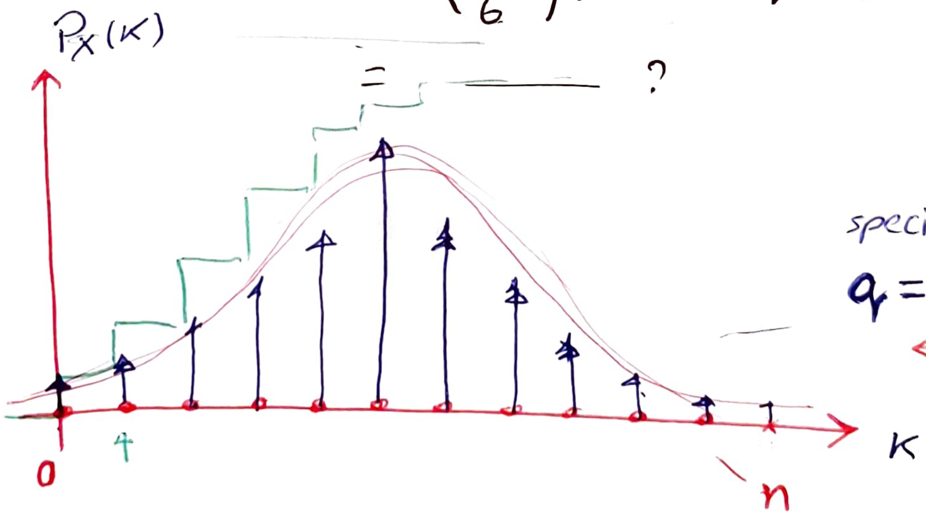
$$= \binom{n}{k} p^{k} (q)^{n-k}$$

$q = 1-p$

P(6 correct questions of 10 questions)

$$= \binom{10}{6} * (0.25)^{6} \times (0.75)^{4}$$

= ______ ?



special case

$q = p = 0.5$

$\bar{X}$ = ? mean = n * p

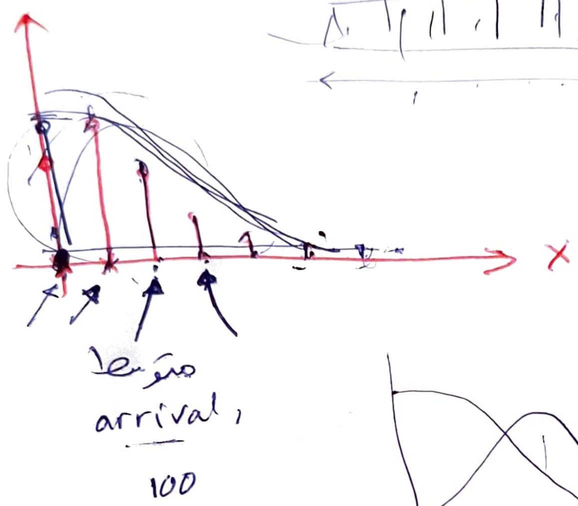
σ_X^2 = ? va = n * p * 1-P

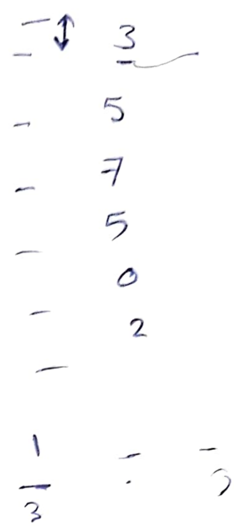
→ Poisson Distribution (Discrete)

→ Poisson Random Variable

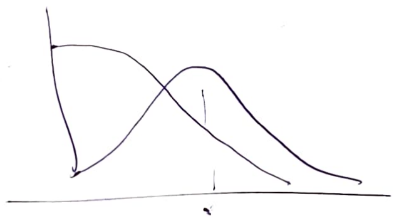
→ Traffic

→ Packet switched network





λ : mean (rate) of arrival



$$\rightarrow \quad P_X(K) = \frac{\lambda^K e^{-\lambda}}{K!}$$

The Poisson distribution is used to model the number of events occurring in a fixed interval of time or space when events happen independently at a constant average rate. Applications include predicting customer arrivals (e.g., at banks or call centers), modeling rare events like earthquakes or accidents, and analyzing failures in manufacturing or systems. It is also used in healthcare (e.g., hospital admissions), telecommunications (e.g., network traffic), and insurance for risk management. In biology and ecology, it models the distribution of organisms. Overall, it is ideal for rare, random, and independent events.