Probability Course

Lecture 3

Created by : Israa Abdelghany
LinkedIn : www.linkedin.com/in/israa-abdelghany-4872b0222
GitHub : https://github.com/IsraaAbdelghany9

Session – 3

⇒ we will discuss :-

- Covariance matrix (statistical tools).
- Discrete Probability distributions (PMF, CDF).
- statistical tools for discrete distribution.
- Intro to continuous distributions.

⇒ Review :-

mean $\quad \bar{x} = \mu_x = \sum_i x_i \, P_x(x_i)$

variance $\quad \sigma_x^2 = \overline{(x_i - \bar{x})^2} = \sum_i (x_i - x)^2 P_x(x_i)$

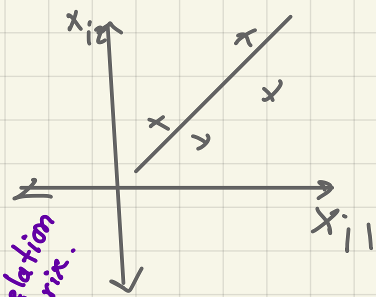
$$= \overline{x^2} - \bar{x}^2$$

$$= \overline{x^2} - \mu_x^2$$

covariance $\quad \sigma_{xy} = \overline{(x_i - \bar{x})(y_i - \bar{y})}$

Covariance Matrix "C" ≡ "Correlation matrix"

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



$$\vec{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix}$$

if more than 2 var.

$$C = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{bmatrix}$$

Correlation matrix.

normalized version of covariance matrix

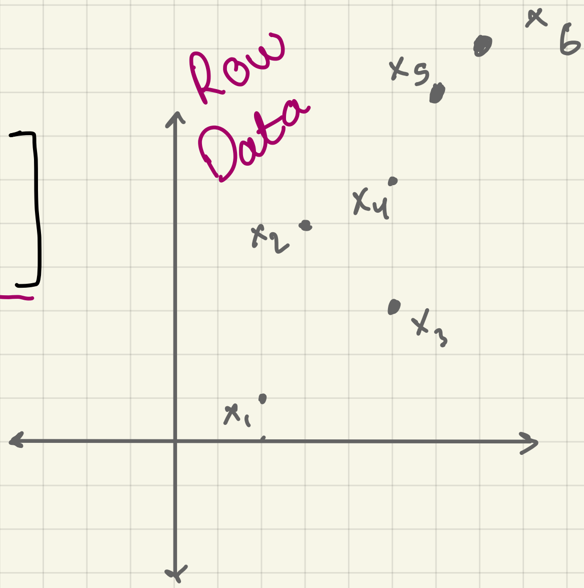
Question :- Calculate the Covariance :-

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

Steps

$\boxed{1}$ ⇒ firstly we can write them in row vector or column vector.

$$x = \underbrace{\begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 5 & 3 & 6 & 7 & 8 \end{bmatrix}}_{\text{can be in column vector representation}}$$



PC 1

PC 2

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

or can be in → Row vector representation

$\boxed{2}$ ⇒ Second step is to Calculate mean for each column in column vector or row in row vector

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

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

$\boxed{3}$ ⇒ third step is to calculate the variance

$$\sigma_\alpha^2 = \frac{(2-4.5)^2 + \cdots\cdots + (7-4.5)^2}{n \text{ or } n-1}$$

6 ← n — if it's the whole set

5 ← $n-1$ — if it's a sample (mean biased)

discussed in statistics

$$\sigma_\alpha^2 = \checkmark$$

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

$\boxed{4}$ Calculate x centered matrix :-

→ 4^{th} step:-

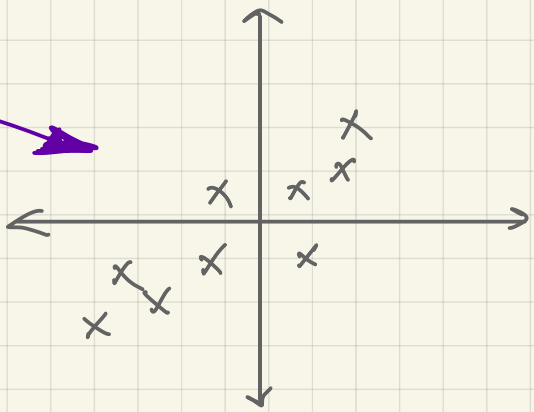
$$x_c = \begin{bmatrix} \alpha_1 - \mu_\alpha & \alpha_2 - \mu_\alpha & \cdots \\ \beta_1 - \mu_\beta & \cdots & \end{bmatrix}$$

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

centered $\underline{\underline{x}}$

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

x-centered matrix representation



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

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

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

number in the given points

mean of the row or column.

to calculate Covariance matrix :-

if data are in column vectors

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

OR $\frac{1}{n-1}$

if data are in row vectors of data matrix

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

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

⇒ to convert it to Correlation matrix ⇒ normalize it.

⇒ divide 1st column by the σ_α and second column by σ_β and 1st row by σ_α and 2nd row by σ_β ⇒ so that diagonal = 1

std

OR simply after calculating x-centered ⇒ normalize it

that is σ_α^2 so the result = 1

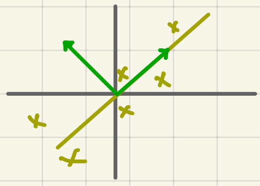
normalized x centered

$$X_n = \begin{bmatrix} \frac{\alpha_i - \mu_\alpha}{\sigma_\alpha} \\ \frac{\beta_i - \mu_\beta}{\sigma_\beta} \end{bmatrix} \Rightarrow \frac{1}{n} X_n X_n^T = \begin{bmatrix} \sum_1 \frac{(\alpha_i - \mu_\alpha)^2}{n \, \sigma_\alpha^2} & \frac{1}{n} \sum_1 \frac{(\alpha_i - \mu_\alpha)(\beta_i - \mu_\beta)}{\sigma_\alpha \, \sigma_\beta} \\ \text{Same} & \frac{1}{n} \sum \frac{(\beta_i - \mu_\beta)^2}{\sigma_\beta^2} \end{bmatrix}$$

$$\therefore R = \begin{bmatrix} 1 & \rho_{xy} \\ \rho_{xy} & 1 \end{bmatrix} \Rightarrow \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y}$$

Covariance matrix aims to get the relation between features

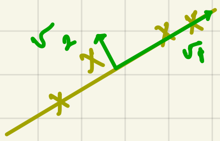
* eigenvectors of the covariance matrix are principle components of data.



→ eigen vectors are orthogonal (prependicular on each other)

- we will have 2 eigen vectors
  - → one represents the direction of data
  - → the other will be in direction with minimum variability

⇒ In the example we have you will find that one of the eigenvalues is much higher than the other one and that's true because data has higher variability in direction and the other one doesn't have that much variability



λ_1 will be higher than λ_2

Covariance matrix answer of the example is:-

$$C = \begin{bmatrix} 3.5 & 4.25 \\ 4.5 & 8 \end{bmatrix} \qquad \vec{v_1} = \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix} \qquad \vec{v_2} = \begin{bmatrix} -0.82 \\ 0.57 \end{bmatrix}$$

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

→ that means that there's info in both directions but λ_1 has much info (the variability is high) but λ_2 is small I'll lose only some info if I ignored it to Simplify the problem.

(Dimensionality reduction

→ All analysis is applied to centered data.

→ PCA → reduce interprability of data

→ Now we know that $\vec{v_1}$ has more variabity so we can define data relative to it

① $\vec{x_1} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

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

③ $\| Proj_{\vec{v_1}} (\vec{x_{1_c}}) \| = \left\| \frac{\vec{x_{1_c}} \cdot \vec{v_1}}{\cancel{\vec{v_1} \cdot \vec{v_1}}} \cancel{\vec{v_1}} \right\| = \begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix} = -4.7$

↳ unit vector = 1

④ So now we can define $\vec{x_{1_c}}$ as $\vec{x_{1_c}} \cong -4.7 \vec{v_1} \cancel{+ \dots \vec{v_2}}$

No Need for $\vec{v_2}$ now

⑤ $\vec{x_{1_c}} \cong -4.7 \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix} = \begin{bmatrix} -2.67 \\ -3.85 \end{bmatrix}$ # which is approximatly $\vec{x_{1_c}}$

So we didn't lose alot info and simplify the data

6 $\vec{x_1} \cong \begin{bmatrix} \text{---} + \mu_\alpha \\ \text{---} + \mu_\beta \end{bmatrix}$ # to get the x_1 from x_{1c} just add μ

SUMMARY

→ PCA : eigen decomposition of covariance matrix C

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

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

- SVD of matrix A ; eigen decomposition of AA^T
- PCA : SVD of X → assignment → instead of

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

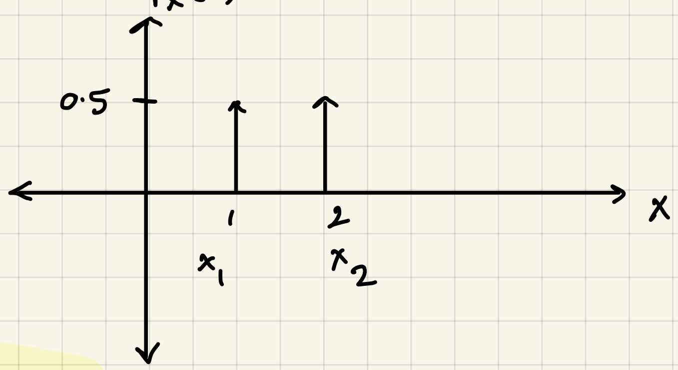
$C = \frac{1}{n} \, SVD \, (X)$

Discrete Probability Distributions

→ Random variable (X)

→ $P_X(x)$

$P_x(x)$ → PMF, Prob.



→ types of distributions:-

- → Uniform distribution ✓
- → Bernoulli distribution ✓
- → Binomial ✓
- → Poisson ✓
- → Geometric
- → hypergeometric

⋮

etc

we will discuss

→ mean, variance of distributions

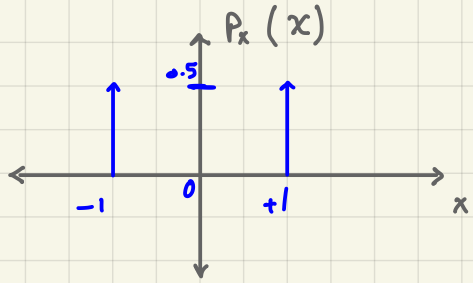
→ CDF

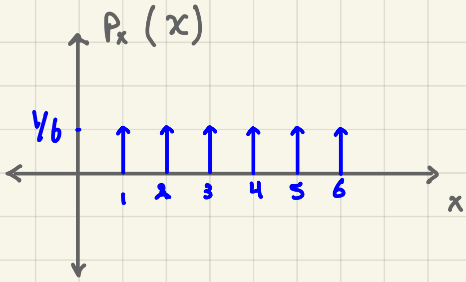
cummulative distribution function

• transition towards discussing continuous distributions

Uniform (discrete) distribution:-

→ All outcomes are equally likely "equiprobable"





→ if z is uniformly distributed R.V.

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

→ this is the simplest and least important probability
it doesn't give much information about data

→ reflect the lake of your knowlage

↳ ex:- probability to answer question you don't know :-

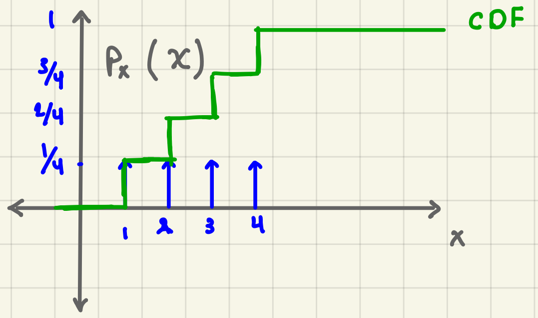
the probability of giving right answer = prob of wrong answer too

So it reflects the lake of knowlage (you don't know)

CDF

احتمالية اختيار الرقم أو رقم أقل منه

CDF : $F_x(x) = P(X \leq x) = P_x(x_i) + P(X < x_i)$



$n = 4$

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

$i = 1:4$ ~ on

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

$F_x(X \leq x_i) = P(X \leq 1.3) = \cancel{P_x(1.3)} + \cancel{P(X < 1.3)} = \text{Zero} + \frac{1}{4} = \frac{1}{4}$ (zero, 1/4)

Properties of CDF :-

→ $F_x(-\infty) = 0$ & $F_x(+\infty) = +1$

→ non decreasing : if $x_2 > x_1 \Rightarrow F_x(x_2) \geq F_x(x_1)$

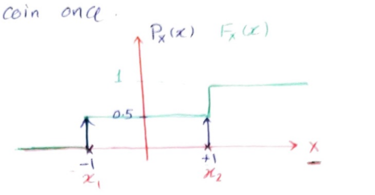
⇒ from CDF we can calculate the probability of specific event by subtract the CDF value after − CDF value before

⇒ one of important parts of CDF is calculating x is in specific range.

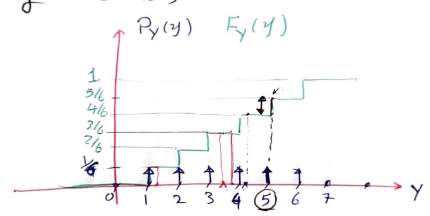
↳ $P(x_1 < X \leq x_2) = F_x(x_2) - F_x(x_1)$

examples

1 → tossing a coin once.



2 → rolling a die once,



$P_x(5) = F_x(5^+) - F_x(5^-) = 5/6 - 4/6 = 1/6$

$P_x(4.2) = F_x(4.2^+) - F_x(4.2^-) = 4/6 - 4/6 = 0$

$P(X \leq 3.5) = F_x(3.5) = 3/6$ → the defination of CDF

1,2,3

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

1 | 2 | 3 | 4 | 5 | 6

2/6

- if x is discrete uniformly distributed R.V.

$\therefore \rightarrow P_x(x) = \begin{cases} 1/n & \text{for } i = 1:n \\ 0 & \text{otherwise} \end{cases}$

$\rightarrow \mu_x = \bar{x} = \sum_{i=1}^{n} x_i \, P_i(x) = \dfrac{\sum_{i=1}^{n} x_i}{n}$

$\Rightarrow \sigma_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \, P_x(x_i)$

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

Example:-

$\rightarrow$ Rolling a dice once.-

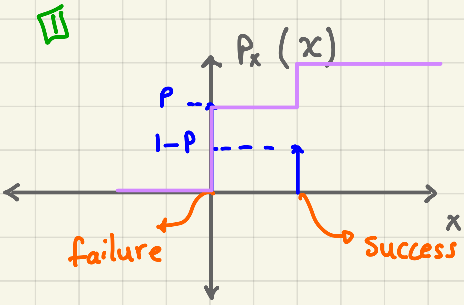
$\bar{x} = 3\,5 = \dfrac{1+2+3+4+5+6}{n}$

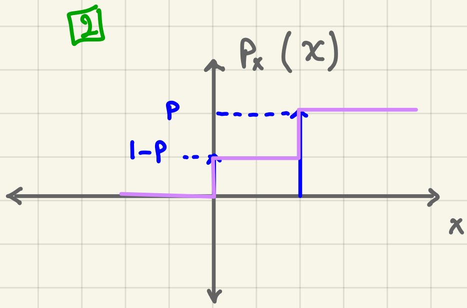
$\sigma_x^2 = \dfrac{1}{6} \sum_{i=1}^{6} (x_i - 3.5)^2$

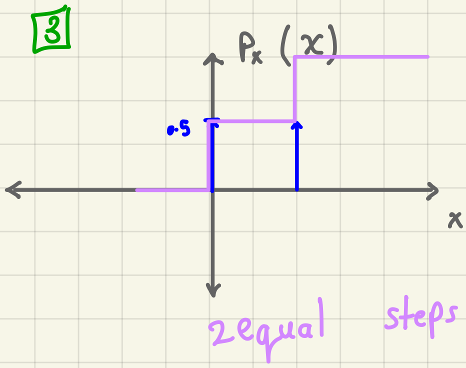
Bernoulli distribution

* second simplist distribution after normal dis.
* has 2 values not one.

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

1


2


3


→ probability has 3 cases as shown above in this example

- failure > success
- 〃 < 〃
- 〃 = 〃

→ $\bar{x} = \sum_i x_i \, P_x(x_i)$

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

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

$p = 1 - q$

$q = 1 - p$

$q + p = 1$

→ $\sigma^2 = \sum_i (x_i - \bar{x})^2 \, P_x(x_i)$

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

$= ?$

OR

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 \longrightarrow p^2 = (1-q)^2$$

$$\overline{x^2} = \sum_i (x_i)^2 \, P_x(x) = \text{Same as the Previous solution}$$

Bionomial Distribution

models:- distribution of number of successes in "n" Bernoulli trials

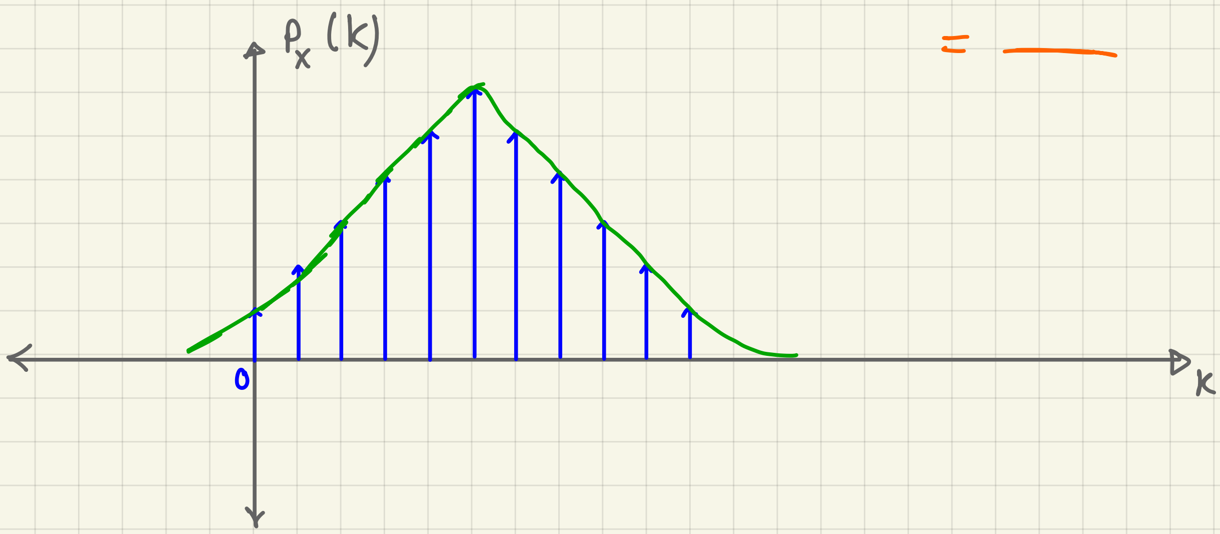
$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}$$

example:- $\boxed{1}$ probability of answering 1 question correct that has 4 choices $= \frac{1}{4}$

$\boxed{2}$ $P(6 \text{ correct of } 10 \text{ questions}) = \binom{10}{6} \times (0.25)^6 \times (0.75)^4$

$= \_\_\_\_$

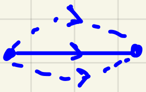
توافيق



Poisson Distribution

→ Poisson Random variable :-

- traffic
- Packet switched network



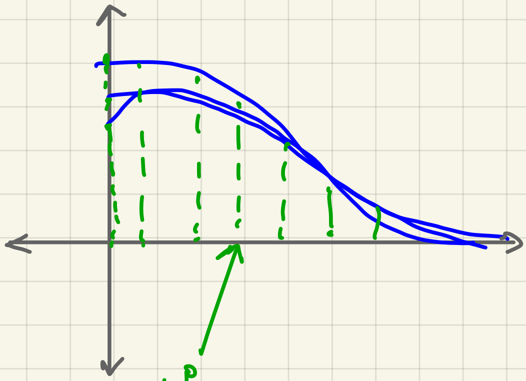
Rules

. Independence

. فترات زمنية متساوية

equal time between them.

λ = mean (rate) of arrival



$$f(k, \lambda) = Pr(X = k) = \frac{\lambda^{k} e^{-\lambda}}{k!}$$

→ How many should I put in the buffer

Customer service أوظف كام حد فى ال

أو اعمل الاشاره مدتها اد ايه

→ Poisson Distribution helps distion making