Probability:

- It's importance in ML:
  1. Uncertainty estimation
  2. Decison making
  3. Algorithm design and analysis.
  4. Model performance evaluation.

- Important topics:
  1- Basic (Conditional probability - Bayes' theorem - Random Variables)
  2- Distributions (Normal - Multinomial - Poisson)
  3- Models: Naive Bayes - logistic regression - Hidden Markov

//

//

Random experiment:

→ is an experment whose outcomes cannot be predicted with certainly.

Sample space: (S)

→ set of all possible outcomes of arandom experiment.

Discrete (countable)

finite

infinite

set {1.....}

Continous

measurements

intervals (range)

$S =]0, 30]$

وحسب الوحدة

Event: (A)

→ Is asubset of sample space.

$A \subset S$

$P(S) = 1$

$P(A) = \frac{N(A)}{N(S)}$

1) $\underset{\text{impossible}}{0} \leq P(A) \leq \underset{\text{Certain}}{1}$

2) $P(S) = 1$

3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4) $P(A^c) = 1 - P(A)$

DeMorgan Laws:

1) $A^c \cap B^c = (A \cup B)^c$

2) $A^c \cup B^c = (A \cap B)^c$

Techniques of Counting:

Multiplication
and
steps

Addition
or/any

Permutation
order
no repetitions (distinct)
selection r
from n
^nP_r

Combination
no order
no repetitions
^nC_r

$$^nP_r = \frac{n!}{(n-r)!}$$

$$^nC_r = \frac{n!}{(n-r)!\, r!}$$

Conditional Probability:

→ The probability of an event $\boxed{B}$ occuring when it is known that ~~the~~ some other event $\boxed{A}$ has occured is called Conditional probability.

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

Intersection

given

?

Intersection/Multiplication Rule:

$$\boxed{P(B \cap A) = P(B/A)\,P(A)}$$

Note:

$\boxed{1}$ if A, B are two disjoint event so

$$P(A \cap B) = 0$$

$\boxed{2}$ if A, B are independant events

→ $P(A/B) = P(A)$ $\quad P(B/A) = P(B$ $\quad P(A \cap B) = P(A).P(B$

Discrete Random Variable:

→ A random Variable is a function that associates a real number with each element in sample space.

→ If the range of a random Variable is Countable set (finite or infinite) the R.V. is said to be discrete R.V.

1 Probability Mass Function (PMF)
Distribution (PDF)

$f(x) = P(X = x)$ denotes the Prob that R.V. X takes the Value of X.

→ PMF must satisfy:

① $f(x) \geq 0$

② $\sum_{R_X} f(x) = 1$

[2] Cumulative Distribution Function (CDF)

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

→ CDF must satisfy:

① non-decreasing function

② $F(-\infty) = 0$

$F : \mathbb{R} \to [0,1]$

③ $F(\infty) = 1$

[3] Expectations $E(x)$ / Mean μ μ_x

X : random Variable

x : Value of X

$f(x)$: PMF/PDF

$$E(X) = \sum x f(x)$$

→ The average Value of R.V. over longrun.

→ The weighted sum of R.V. Values

Properties of Expectation:

① $E(g(x)) = \sum g(x) f(x)$

② $E(ax+b) = aE(x) + b$

③ $E(c) = c \longrightarrow c$ is Constant

④ $E(ax \pm bY) = E(ax) \pm E(bY) = aE(x) \pm bE(Y)$

$\boxed{4}$ Variance $V(x) = \sigma^2$

Standard deviation $\sigma = \sqrt{V(x)}$

$V(x) = E(x-\mu)^2$

$= \sum (x-\mu)^2 f(x)$

$V(x) = E(x^2) - (E(x))^2 = E(x^2) - \mu^2$

Properties of Variance:

① $V(ax+b) = a^2 V(x)$

② $V(c) = 0$

③ $V(ax \pm bY) = a^2 V(x) + b^2 V(Y)$

$x, y \rightarrow$ R.V

5 Moment Generating Function (MGF)

→ The mean of random variable is called 1st moment.

In general $\boxed{r^{th} \text{ moment} = E(x^r)}$

→ MGF is a function that can be used to derive various moments of R.V.

$$M_x(t) = E(e^{tx}) = \sum e^{tx} f(x)$$

$$E(x) = \frac{d}{dt} M_x(t)\Big|_{t=0} = \text{Mean} = 1^{st} \text{ moment}$$

$$E(x^2) = \frac{d^2}{dt^2} M_x(t)\Big|_{t=0} = 2^{nd} \text{ moment}$$

Special discrete Prob. distributions:

[1] Bernoulli Trials:

→ a random experiment with $\boxed{\text{two possible}}$ outcomes that may be success or failure.

→ PMF of Bernoulli Trials: $f(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$

→ Mean (μ): $\sum x f(x) = p$

→ Variance: $E(x^2) - \mu^2 = p(1-p)$

[2] Binomial experiment:

→ Consists of $\boxed{n}$ Bernoulli Trials such that:

1. trials are independent.
2. each trial result in only two possible outcomes.
3. Probability of success in each trial denotes as p.

→ PMF: $f(x) = C^n_x \, p^x \, q^{n-x}$, $x \rightarrow$ no. successes

→ Mean μ $E(x) = np$

→ Variance $V(x) \Rightarrow \sigma^2 = npq$

3 Geometric Distributions:

→ As binomial experiment but the trials will be repeated until first success.

P: Prob. of success

$X \quad Geo(P)$ $\quad x$: No. trials until first success

→ PMF: $f(x) = P\underbrace{(1-P)}_{q}^{x-1}$

→ Mean: $E(X) = \frac{1}{p}$

→ Variance: $V(X) = \frac{q}{p^2}$

4 Poisson Distribution:

→ It gives the probability of an event happening a certain number of times (K) within a given interval of times or space.

→ PMF: $f(X) = \frac{e^{-\lambda} \lambda^{X}}{X!}$

λ: average $\frac{\text{rate}}{\text{No.}}$ of success in a given interval

x: No. of success. in given interval.

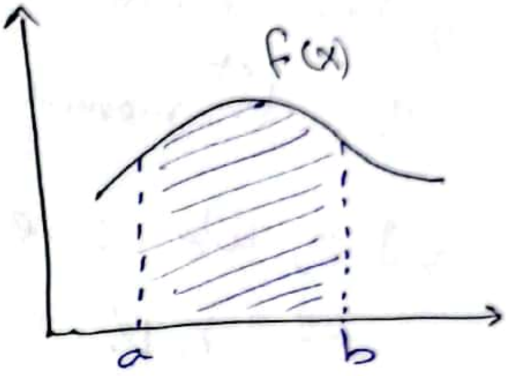
→ $E(X) = \lambda$

→ $V(X) = \lambda$

Continuous Random Variables:

→ When r.v. denoting range of values like lengths, time, distance...

[1] Probability density function (PDF):



1. $f(x) \geq 0$
2. $P(a < x < b) = \int_a^b f(x)\,dx$
3. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

∞ → end

$-\infty$ → start

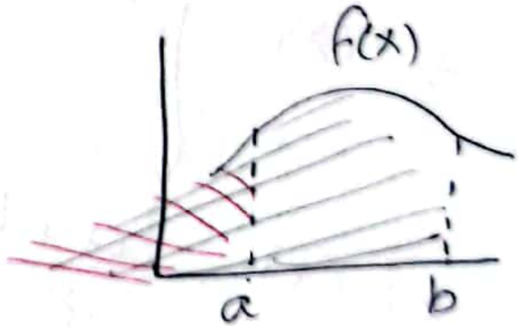
[2] Cumulative distribution function (CDF):

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(\ )\,d$

$-\infty$ → start

→ If given $F(x) = P(X \leq x)$

- $P(x < a) = F(a)$
- $P(x > a) = 1 - F(a)$
- $P(a < x < b) = F(b) - F(a)$



Moment Generating Function: MGF

$$E(x^r) = \frac{d^r}{dt^r} M_x(t) \Big]_{t=0} \longrightarrow r^{th} \text{ moment}$$

→ The mean of a r.v. is referred to as the 1^{st} moment of prob. distribution

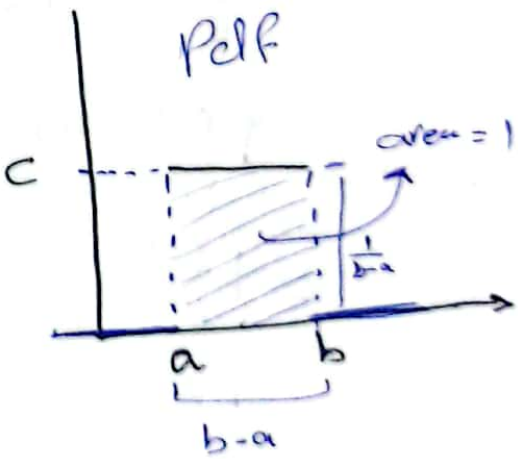
→ So, we use MGF to derive various moment of r.v..

Special Continuous Prob. Distr.

$\boxed{1}$ Uniform distribution:-

→ A r.v. that is Uniformly distributed over an interval $[a, b]$ has constant pdf

→ $f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & o.w \end{cases}$



→ $E(x) = \frac{a+b}{2}$

→ $V(x) = \frac{(a-b)^2}{12}$

2 Exponential Distr. :

- if $x \sim$ poisson (λ)

x : no. of events per Unit of time

→ $T \sim \exp(\lambda)$

T : - The time required for 1^{st} event to occur
- the time to 1^{st} poisson event occur
- Inter arrival time between succesive poisson events.

- $F(t) = 1 - e^{-\lambda t}$ ⟸ CDF

- $E(t) = \frac{1}{\lambda}$ $\qquad f(t) =$

- $V(t) = \frac{1}{\lambda^2}$

3 Erlang-k Distribution :

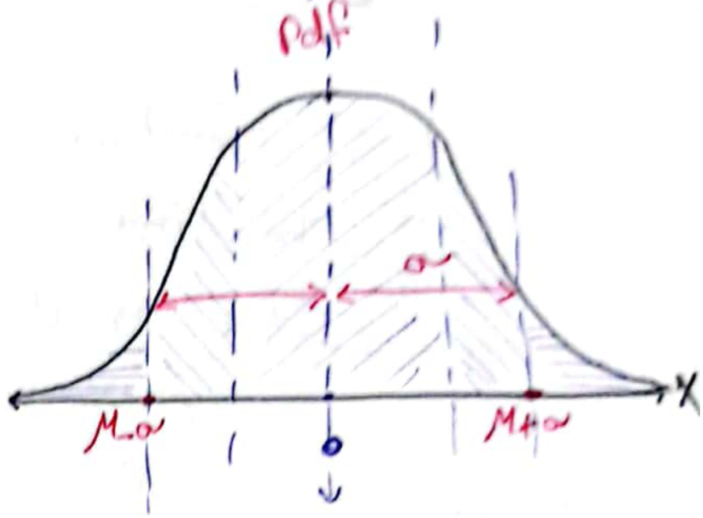
→ It is a generalization of exp. distr. that describe the length of an interval until k event occur in a poisson process or time until k^{th} poisson event.

- $f(t) = \dfrac{\lambda^k e^{-\lambda t} t^{k-1}}{(k-1)!}$, $t > 0$, $k = 1, 2, 3. \ldots$

- $E(t) = \frac{k}{\lambda}$, $V(t) = \frac{k}{\lambda^2}$

4 Normal (Gaussian) distribution:

→ A probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than the data far from mean.



→ $X \sim Normal(\mu, \sigma^2)$

→ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ pdf

→ $f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}}$ standard normal → $-\infty < x < \infty$

↳ at $\mu = 0$ and $\sigma = 1$

→ In Standard normal:

→ $Z \sim N(0,1)$ $\quad \phi(-Z) = 1 - \phi(Z)$

→ $P(-3 < Z < 3) = 1$

→ $\phi(0) = \frac{1}{2}$ $\quad \phi(3) = 1$

$\phi(-3) = 0$

→ $\phi(\infty) = 1$

→ $\phi(-\infty) = 0$

Standarization

$X \longrightarrow Z$

$Z = \frac{X - \mu}{\sigma}$

Discrete joint Probability distribution:

[1] Joint probability mass function:

→ If X, Y are two discrete random Variables the probability distr. for their simultaneous occurrence can be represented by a function $f_{xy}(x,y)$ called joint PMF or joint PDF.

→ $f_{xy}(x,y) = f_{xy}(X=x, Y=y)$

① → $0 \leqslant f_{xy}(x,y) \leqslant 1$

② → $\sum_{x}\sum_{y} f_{xy}(x,y) = \boxed{1}$

⇒ check dependency:

* X, Y are indep. if

$$\underbrace{f_{xy}(x,y)}_{\text{joint}} = \underbrace{f_x(x)}_{\text{marginal of } x} \cdot \underbrace{f_y(y)}_{\text{marginal of } y} \rightarrow \text{for all } \underline{x,y}$$

⇒ Conditional probability:

$$f_{y/x} = \frac{f_{xy}(x,y)}{f_x(x)} \quad \begin{matrix} \rightarrow \text{Joint} \\ \rightarrow \text{marginal of given} \end{matrix}$$

Covariance and Correlation Coefficient:

$$Cov(x,y) = E\big((x-\mu_x)(y-\mu_y)\big) = \sum \frac{(x-\mu_x)(y-\mu_y)}{n-1}$$

σ_{xy}

$0 < Cov \Rightarrow$ +ve relationship $x\uparrow \; y\uparrow$

$0 > Cov \Rightarrow$ -ve relationship $x\uparrow \; y\downarrow$

$0 = Cov \Rightarrow$ no relationship

note:

→ we can't know by Cov How strong the relationship.

Correlation Coefficient:

$$\rho(x,y) = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y}$$

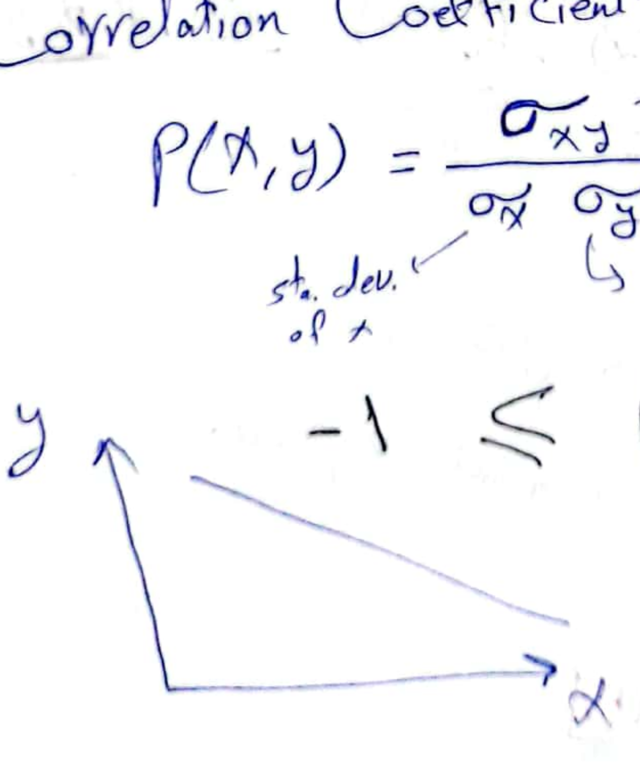
Cov(x,y)

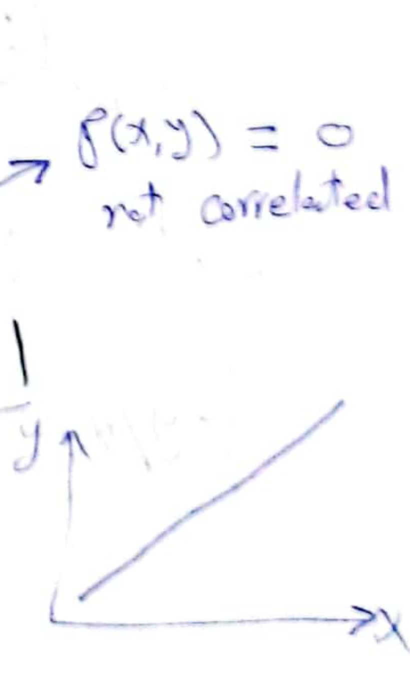
sta. dev. of x

stan. devi. of y

$$-1 \le \rho(x,y) \le 1$$

$\rho(x,y) = 0$ not correlated





If X, Y are two random variables whose joint pdf $f_{xy}(x,y)$

Discrete

$\rightarrow \mu_x : E(x) = \sum x \, f(x)$ → margin of x

$\rightarrow \mu_y : E(y) = \sum y \, f(y)$ → margin of y

$\rightarrow E(x,y) = \sum_{x,y} xy \, f(x,y)$

$$\Rightarrow \boxed{Cov(x,y) = E(x,y) - \mu_x \mu_y}$$

In Continuous :

$$f_x(x) = \int_y f_{xy}(x,y) \, dy$$ → vertical strip

$$f_y(y) = \int_x f_{xy}(x,y) \, dx$$ → horizontal strip

$\rightarrow X, Y$ are independent if $f_{xy}(x,y) = f_x(x) \cdot f_y(y)$