Probability

→ measures the likelihood of any event happening.

baised coin (H H)

unbaised coin (HT) → $\boxed{\frac{1}{2}}$ ⇒ 50%.

HT

→ 1000 → 700 H, 300 T

$\frac{700}{1000} = \frac{7}{10}$

$\frac{300}{1000} = \frac{3}{10}$

Sample Space : The set of all possible outcame

→ unique/Distinct

{H, H} ✗ {H, T}.

Event A subset of all possible outcames.

Event ⊆ Sample space

Rolling a Die.

Sample Space {1, 2, 3, 4, 5, 6}

not

for single Dice roll. 1/6 $\left(\frac{2}{6}\right)$

two dices

1st 2nd

{(1, 1), (1, 2), (1, 3)

1st die 2nd dice

36 Events

$\left(\frac{1}{36}\right)$

Types of probability

→ classical : Based on equally likely outcome.

→ Empirical :- (Based on observation or experiments

→ scenrio Based

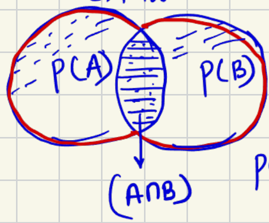
→ Montecarlo Methods.

Subjective to the personal judgement or experience

1) Addition Rule

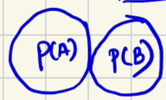
$P(A) \rightarrow$ prob

Set rule $\hookrightarrow$ event



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

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



$P(A \cap B) = 0$

independent

2) <u>Multiplication Rule</u>

$$P(A \cap B) = P(A) * P(B)$$ <independent event>

3) Complementary Rule

$$P(\sim A) = 1 - P(A)$$

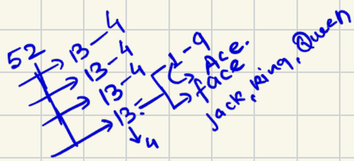
not rolling a 3 on die

$$1 - \frac{1}{6} = \frac{5}{6}$$

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * \cancel{P(B)}}{\cancel{P(B)}}$$

$$= P(A)$$



Prob. of drawing a ace given that you <u>drew a face card</u>

4 Ace   <u>16 face card</u>

$$\frac{4}{16} = \frac{1}{4} = 25\%.$$

Independent v

→ outcome of one event does not affect the outcome of another

dependent

out come of one event affects the outcome of another.

Drawing cards from a deck (without replacement)

$\frac{1}{52} \times \frac{1}{51} \times \frac{1}{50}$ (without replace

$\frac{1}{52} \times \frac{1}{52} \times \frac{1}{52}$ (with repacement)

Rolling a die to get even numbers

$(1, \underline{2}, 3, \underline{4}, 5, \underline{6})$

$3/6 = 1/2$

There are 3 red 2 blue mables.

if you peak a mable and do not repace it. what is the prob. of picking a red marble first and then a blue marble

$\{R, R, R, B, B\}$

$\frac{3}{5} \times \frac{2}{4}$

Bayes' Theorem

find the prob. of an event given the prob. another related event

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$= \frac{\cancel{P(B)} \times P(A)}{\cancel{P(B)}}$$

$$= P(A)$$

A factory has two machine producing parts

Machine A ← 60% of parts

Machin B ← 40% of parts

1% of the parts from A are. defective

2% from B.

if a part is found to be defective what is the prob. it come from Machine A

$P(A) = 0.6$
$P(B) = 0.4$
$P(D|A) = 0.01$
$P(D|B) = 0.02$

$$P(A|D) = \frac{P(D|A) * P(A)}{P(D)}$$

Low of total prob

$$P(D) = P(D|A) \cdot P(A) + P(D|B) * P(B)$$
$$= 0.01 \times 0.6 + 0.02 \times 0.4$$
$$= 0.006 + 0.008$$
$$= 0.014$$

$$\frac{0.01 * 0.6}{0.014} = \frac{0.006}{0.014} = \frac{6}{14} = \frac{3}{7}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{— (1)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{— (2)}$$

$$P(A \cap B) = P(B|A) * P(A) = P(A|B) * P(B)$$

$$\Rightarrow P(B|A) * P(A) = P(A|B) * P(B)$$

$$\boxed{\Rightarrow P(A|B) = \frac{P(B|A) * P(A)}{P(B)}}$$

- → Naive Bayes Classifier
- → Bayessian Networks
- → Bayesian Inference
- → Regularization & Model selection
- → Bayesian opt.
- → Markov chain

Random variable

$\boxed{\text{random process}}$

A quantity whose $\boxed{\text{random values}}$ by the outcome of a random process.

→ Depends on the chance & uncertainty involve in an experiment or situation

i) Discrete Random variable.

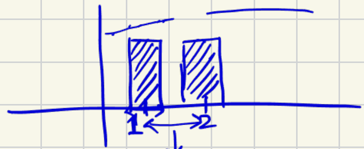
↦ it can only on a Countable number of values

Roll a diece $\{1, 2, 3, 4, 5, 6\}$

$\{1 - 2\}$

$\langle 1.1, 1.2\ 1.3\ 1.4 \rangle$

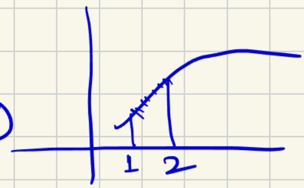
coin



Sample Space.

ii) Continuous random variable

height (continuous R.V.)
temperc



Discrete $\underbrace{X}_{R.V.} : \underbrace{S \to \{x_1, x_2, x_3 \cdots x_n\}}_{\text{Sample space}} \leftarrow \boxed{1.}$

continous $\underbrace{X}_{R.V} : S \to \mathbb{R}$

$S \to [1, 2]$

Prob. Mass function

provide prob. that a $\boxed{\text{decrete R.V.}}$

$$\boxed{P(X=x)} = p(x)$$

PMF

Discrete Random variable

value.

$$\boxed{\sum_{X \in X} P(x) = 1}$$ Non-negative value.

Expected value Calculation (Mean)

$$E(X) = \sum_{x \in X} \underline{x \cdot P(X=x)}$$

$[0, 1]$

0.8   0.~~7~~

0   Mean   1

$0 \times 0.5 + 1 \times 0.5 = 0.5$

0.7

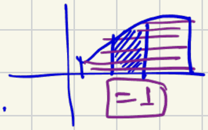
Variance.

$$Var(X) = \sum_{x \in X} (x - E(X))^2 \cdot P(X=x)$$

$$Std(X) = \sqrt{Var(X)}$$

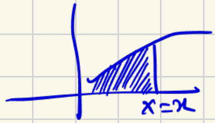
Continuous R.V.

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$



$$\boxed{\int_{-\infty}^{\infty} f(x)\,dx = 1}$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$



Cumulative Distribution function

$$F(x) = P(X \leq x)$$

$$\boxed{\sum_i P(X = x_i) = 1}$$

chain rule of conditional prob.

$$P(A \cap B) = \boxed{P(A|B) \cdot P(B)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \leftarrow \text{join prob. } A \,\&\, B.$$

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \cdots \cap A_n) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots A_{n-1})$$

$$P(A_1 \cap A_2 \cap A_3) = \underline{\underline{P(A_1)\,P(A_2|A_1)}\;P(A_3|A_1 \cap A_2)}$$

$$\underline{P(A_1 \cap A_2)\;P(A_3|A_1 \cap A_2)}$$

$$P(A_1 \cap A_2 \cap A_3)$$

Three consecutive exams.

M P C

$$P(M) = 0.8$$
$$P(P|M) = 0.7$$
$$P(C|M \cap P) = 0.9.$$

$$P(M \cap P \cap C) = \boxed{P(P)\,P(M|P)}\,P(C|M \cap P)$$

$$= P(M)\,P(P|M)\,P(C|M \cap P)$$

$$= 0.8 * 0.7 * 0.9$$

$$= 0.504 \simeq 50.4\%.$$

1) <u>Bernauli Distribution</u>

A R.V. X is said to follow B.D. if it takes the value 1 with prob. P and the value 0 with $(1-P)$

<u>PMF</u>

$$P(X=x) = \begin{cases} P & \text{if } x=1 \\ 1-P & \text{if } x=0 \end{cases}$$

$$1 \cdot P + (1-P) \cdot 0$$

$$\boxed{E(X) = P} \qquad \text{var} = P(1-P)$$

Skewness of B.D.

if $p = 0.5$, the distribution is symmetric

if $\underline{p > 0.5}$ the distribution is $\boxed{\text{-vely skewed}}$ More 1

if $\underline{p < 0.5}$ the distribution is $\boxed{\text{+vely skewed}}$ More 0.

→ Suppose we flip a fair coin 1

$$\left.\begin{aligned} P(\underline{\text{heads (success)}}) &= 0.5 \\ P(\text{tails (failure)}) &= 0.5 \end{aligned}\right\} \text{symmetric}$$

1 0

$$\left.\begin{matrix} 0.4 \\ 0.6 \end{matrix}\right\} \text{+ve skewed}$$

$$\left.\begin{matrix} 0.6 \\ 0.4 \end{matrix}\right\} \text{-vely skewed}$$

$$\left.\begin{aligned} p_{spam} &= 0.50 \\ p_{implotant} &= 0.30 \\ p_{normal} &= 0.20 \end{aligned}\right\} 1$$

$p(\text{classified as spam} \mid \text{spam}) = 0.90$

$p(\text{classified as spam} \mid \text{Implant}) = 0.10$

$p(\text{classified as spam} \mid \text{normal}) = ?$

$p(\text{classfied as normal} \mid \text{nonal}) = 0.9$

$p(\text{classfied as Importan} \mid \text{implant}) = 0.85$

$p(\text{classified as spam} \mid \text{spam}) = 0.9$

$p(\text{classfied as spam} \mid n) = 1 - \cancel{p(CAN \mid N)} - \boxed{p(CAI \mid N)}$

Geometric Random variable

→ The number of trails (idependent) required to get the first Succes in a series of Bernauli trails

$$\text{Bernauli trails} \to 1 \; (p), \quad 0 \; (1-p)$$

K^{th} trails are Success. ⇒ $(K-1)$ times trails are fail.

are Success ↓ Beranauli trails.

prob. Mass fn.

$$p(X = \underset{\downarrow}{k}) = (1-p)^{k-1} p$$

[discrete]

$$E(X) = \frac{1}{p} \qquad Var = \frac{1-p}{p^2}$$

Suppose you are rolling a fair die
→ How many rolls it will take to roll a 6 for first time
→ prob. that you roll a 6 for first tim on the third roll

$$\boxed{X=3}$$

$$p = \left(1-\frac{1}{6}\right)^{3-1}\left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^2\frac{1}{6}$$

$$= 0.1157$$
$$= 11.57\%$$

1st head on the forth flip.

$$P(X=4) = \left(1-\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = \frac{1}{16}$$

$$= 0.0625 \quad = 6.25\%$$

Poisson Random Variable

→ counts the number of events occurring within a fixed interval of time or space.

Condn
→ All events occur independently of each other
→ Avg. rate (λ) of occurance is const.
→ Two events cannot occur at exactly same time

PMF
$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

X ↓ number of events occuring in a fixed interval

λ ← avg. number of events
k ← actual number of occurance of event

$$E(X) = \lambda$$
$$Var(X) = \lambda.$$

5 emails per hrs.

$X \leftarrow$ number of emails you recieve in a given hr.

prob. of recieving 3 emails in the next hr.

$x = 3$

$\lambda = 5$

$$p(x=3) = \frac{5^3 e^{-5}}{3!} = 0.1404$$

$=$

$\lambda = 0$	$X = 15$
$\lambda = 2$	$\lambda = 10$

$\lambda = 4$

prob. of obseving fewer than 2 vistors in hour.

$$p(x<2) = p(x=0) + p(x=1)$$

defective $2/hr \rightarrow 3$ hrs. / rate

$\rightarrow 2 \times 3$

$= 6 = \lambda.$

$$p(x>2) = 1 - p(x \leq 2)$$

$$= 1 - p(x=0) - p(x=1) - p(x=2)$$

Binomial random variable.

$\rightarrow$ Counts the number of Success in a fixed number of independent B.T.

$$X \sim \text{Binomial}(n, p)$$

$n =$ number of trials

$p =$ prob. of Success in each trails

$q = (1-p) =$ prob. of failure

PMF

$$p(X=k) = \boxed{\binom{n}{k}} p^k (1-p)^{n-k}$$

1 — S
0 — F

$\boxed{011}$ $\boxed{110}$

Mean (Expected value) $= E(X) = np$.
$\mathrm{var}(X) = np(1-p) = npq$.

if you flip a coin 10 times $(n=10)$ $P(H) \to$ Success $= 0.5$

$$X \sim \text{Binomial}(10, 0.5)$$

$\boxed{X=4}$

$$p(X=4) = \binom{10}{4} (0.5)^{4} (0.5)^{6}$$

10000
2

$= 0.205 = 20\%$

Approximating a Binomial R.V. with poisson

$$X \sim \text{Binomial}(n, p)$$

$n \leftarrow$ large. low success. $p =$ is small

$$p(X=k) = \boxed{\binom{n}{k}} p^k (1-p)^{n-k}$$

Stirling's approximation for larger n

$$\leftarrow \binom{n}{k} \simeq \frac{n^k}{k!}$$

$$= \frac{n^k}{k!} p^k (1-p)^{n-k}$$

$$= \frac{(np)^k}{k!} (1-p)^{n-k}$$

mean $= np = \lambda$

$$\underbrace{(1-p)^n}_{\text{large}} \simeq e^{-np} = e^{-\lambda}.$$

$$= \frac{\lambda^k}{k!} e^{-\lambda}$$

1000 light bulb with $p = 0.0001$ being defective

$$\underset{\substack{\downarrow \\ \text{number of bulb are defective}}}{p(x = 2)} = \binom{1000}{2} p^2 (1-p)^{998}$$

$np = \lambda = 1$

$1000 * 0.0001$

$$\frac{1^2 e^{-1}}{2!} = 0.1839$$

18.39%

Uniform R.V

$$f(x) = \begin{cases} 1 & , 0 < x < 10 \\ 0 & , \text{Otherwise} \end{cases}$$

x is uniformly distributed over (0 10)

CDF $\boxed{P(X < 3) = \int_0^3 \frac{dx}{10} = \frac{3}{10}}$

Exponential Random dis.

→ continuous prob. dis.

time or distance events (poisson process)

independent const avg. rate

pdf $f(x;\lambda) = \begin{cases} \boxed{\lambda e^{-\lambda x}} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

Mean $= E[X] = \lambda$

$Var = 1/\lambda^2$

Memoryless property:

$p(X > s+t \mid X > s) = p(X > t)$

$\lambda = 3$ calls/hr.

$E[X] = \frac{1}{\lambda} = \frac{1}{3} = 20$ min.

prob the next call arrives within the next 10 min

$\frac{10}{60} = 0.1667$

$p(X \leq 0.1667) = 1 - e^{-\lambda x}$

$= 1 - e^{-3 * 0.1667}$

$= 0.3935$

39.35%

$p(X \geq y) = 1 - p(X < y)$

prob. of an event occurring in the future independent of how much time has already elapsed

$p(X > 3 \mid X > 2)$

$P(x > 2+1 \mid x > 2)$

$P(x > 1)$

$1 - P(x < 1)$

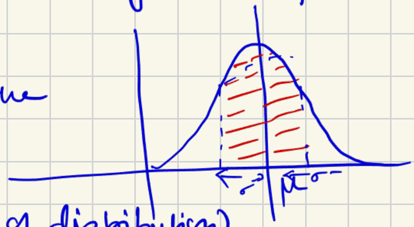
Normal distribution (Gaussian Distribution)

→ Continuous R.V.

→ clusters around a mean (average) expected.

→ Most values are close to the mean

→ fewer values appear as you move from the mean



Mean → center of the distribution

S.d → control the spread (width of distribution)

S.D = 1 (σ) → 68% of data lies with 1σ of the mean ($\mu \pm 1\sigma$)

SD = 2 → 95% of data with 2σ

SD = 3 → 99.7% - - - - - - - 3σ

PDF

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

Standard Normal Distribution

$\mu = 0$

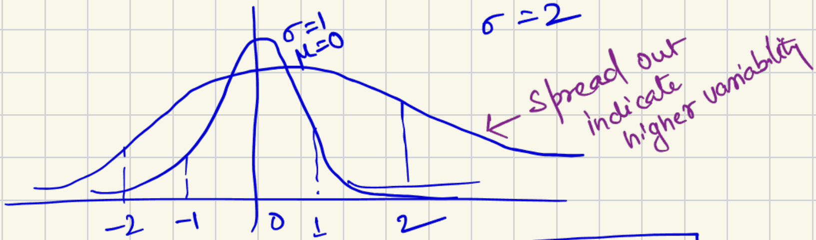
$S.D(\sigma) = 1$

$Z \sim N(0, 1)$

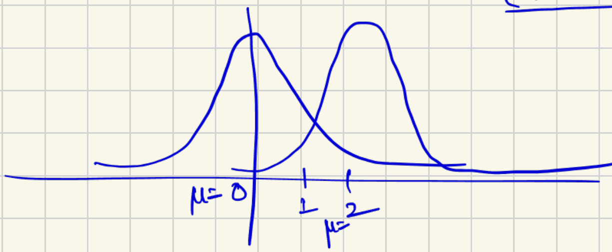
$$Z = \frac{X - \mu}{\sigma}$$

$$PDF = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$\sigma = 1 \quad \sigma = 2 \quad$ Mean $= 0$



$\boxed{\text{Mean} = 2}$



$$\boxed{Z = \frac{x - \mu}{\sigma}} \leftarrow \text{Z-score}$$

Z-table provides the cummulative probabity from the left tail of the Standard Normal Distribution

Prob. the value is less than 1.23.

$\boxed{\text{SND}} \rightarrow \mu = 0$
$\sigma = 1$

Z-table

$Z = 1.23$
$= 0.8907$

The height of Students are normally distributed
↓ C.R.V

mean = 170cm
S.D = 6cm

b/w 164 cm and 176 cm
Z_1 Z_2

$p(164 < x < 176)$

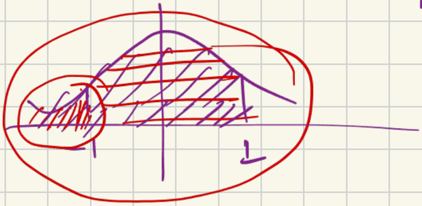
$$z = \frac{x - \mu}{\sigma}$$

$$z_1 = \frac{164 - 170}{6} = -1$$

$$z_2 = \frac{176 - 170}{6} = 1$$

$p(z_1)$

$p(X < z_1) -$



$p(-1 < X < 1)$

$p(X < 1$

$p(Z=1) - p(Z=-1)$

$0.8413 - 0.1587 =$

Parametric Density Estimation $\quad P(X, \theta)$

→ Maximum likelihood Estimation (MLE)

Estimates the parameter of the assumed distribution by maximizing the likelihood fn.

MLE for Bernauli Distribution

10 tosses

(H T H H T H H T T H) ← independent

Estimate p of getting heads using MLE

Number of head = 6

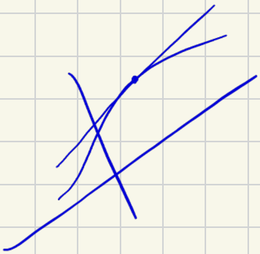
Toss = 10

MLE for Bernauli distrib

$$\hat{p} = \frac{6}{10} = 0.6$$

parameter θ. that maximizes the likelihood fn $L(\theta; x)$ where X represents the observed data.

$$L(\theta; x) = \prod_{i=1}^{n} \underbrace{f(x_i; \theta)}_{\text{PDF}}$$



$$\log(L(\theta; x) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

$$\frac{d}{d\theta} \log(L(\theta; x)) = \frac{d}{d\theta} \sum_{i=1}^{n} \log f(x_i; \theta)$$

$$= \sum_{i=1}^{n} \frac{d}{d\theta} \log f(x_i; \theta)$$

$$\Rightarrow \sum_{i=1}^{n} \frac{d}{d\theta} f(x_i; \theta) = 0$$

$$\Rightarrow \log L(\theta; x) = 0$$

Method of Moments Estimation (MME)

parameter of a prob. distribution by matching sample moments

→ are quantities that describe certain aspects of shape of P.D.

r^{th} moment of R.V. X is defined as $E(X^r)$

1st moment $E(X)$ = mean
2nd moment $E(X^2)$ = variance

Let us $\boxed{X_1, X_2 \text{ ---- } X_n}$ are independent & identical distributed R.V.

$N(\mu, \sigma^2)$. We want to estimate μ and σ

$\theta_1 = \mu \quad \theta_2 = \sigma$

$$E(X) = \mu.$$
$$E(X^2) = \mu^2 + \sigma^2$$
Theoritical moments

Sample based

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$
$$M_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

$$\mu = M_1$$

$$E|X^2| = \mu^2 + \sigma^2 = M_2$$

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

$$\sigma^2 = M_2 - \mu^2$$
$$= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)^2$$

MME $\quad B(n, p) \quad$ n is known

$$p(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E|X| = np$$
$$\text{var} = npq = np(1-p)$$

$$X_1, X_2 \ldots X_k$$

$$M_1 = \frac{1}{k} \sum_{i=1}^{k} x_i$$

$$\boxed{P = \frac{M_1}{n} = \frac{1}{n} \frac{1}{k} \sum_{i=1}^{k} x_i}$$

poision distribution

$$E|x| = \lambda$$
$$E|x^2| = \lambda.$$

$$p(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$x_1 \quad x_2 \quad \cdots \quad x_n$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\lambda = \bar{x}$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$