Probability

Sample space : It is the set of all possible outcomes

Event : An event is a subset of sample space

$$P(A/B) = \frac{P(A \text{ and } B)}{P(B)}$$

$\rightarrow$ Joint P

$\rightarrow$ Marginal P

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

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad - (1)$$

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

from (1) $P(A \cap B) = P(B)\,P(A/B)$

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

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

$$E[x] = \sum_{i=1}^{n} x_i P(X = x_i)$$

$$= \mu$$

Mean

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

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

$$= E[x^2] - (E[x])^2$$

Variance

$$std(X) = \sqrt{Var(x)}$$ } Standard Deviation

Binomial Distribution

$$P(x) = {}^{n}C_{x} \, p^{x} q^{n-x}$$

- $n \rightarrow$ No: of times experiment is performed
- $x \rightarrow$ No: of favourable outcomes
- $p \rightarrow$ Probability of favourable outcome
- $q \Rightarrow$ Probability of unfavourable outcome

$$q = (1-p)$$

$P(x > n) = q^n$ $\qquad$ $P(x \geq n) = q^{n-1}$

$P(x < n) = 1 - q^n$ $\qquad$ $P(x \geq n) = 1 - q^{n-1}$

$$\text{Var} = \frac{1}{p}\left(\frac{1}{p} - 1\right)$$

$$= \mu(\mu - 1)$$

$$\boxed{\mu = \frac{1}{p}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{p}\left(\frac{1}{p} - 1\right)}$$

$$= \sqrt{\frac{1}{p}\left(\frac{1-p}{p}\right)}$$

$$= \frac{\sqrt{1-p}}{p}$$

$$= \frac{\sqrt{q}}{p}$$

Poisson Distribution

$$P(x=n) = \frac{\mu^n e^{-\mu}}{n!} \quad \text{or} \quad = \frac{\lambda^n e^{-\lambda}}{n!}$$

$$\mu = np \qquad \sigma^2 = np \qquad \sigma = \sqrt{np}$$

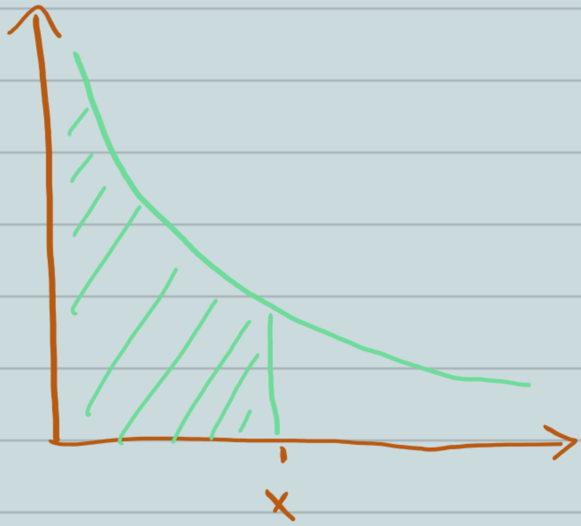
$$= \sqrt{\lambda}$$

$$P(X > n) = 1 - e^{-r}\left[\sum_{x=0}^{n} \frac{r^n}{n!}\right]$$

$$P(x \geq n) = e^{-r}\left[\sum_{x=0}^{n} \frac{r^n}{n!}\right]$$

Cumulative Distribution function

suppose consider the graph below



CDF determines the area under the curve and to the left of point x.

This area is equal to the probability

$P(X \leq x)$

Normal Distribution

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

For the normal distribution PDF

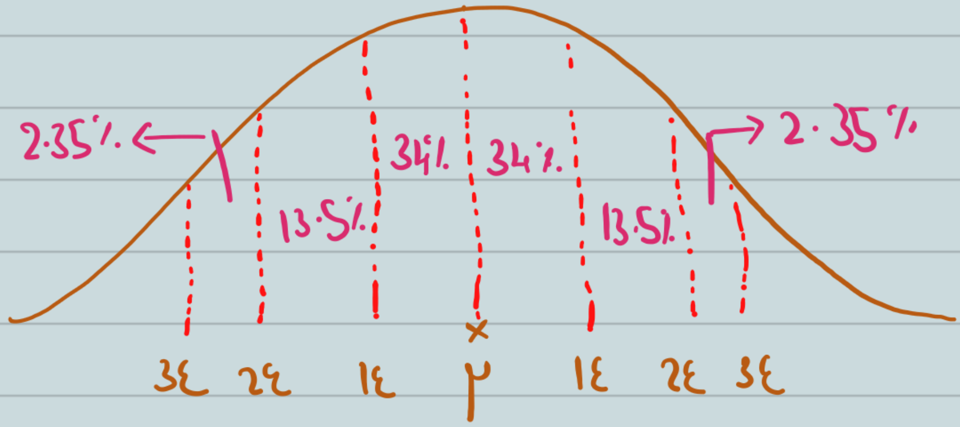
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Emperical rule :

68% of x values $\in [\mu - \varepsilon, \mu + \varepsilon]$

95% of x values $\in [\mu - 2\varepsilon, \mu + 2\varepsilon]$

99.7% of x values $\in [\mu - 3\varepsilon, \mu + 3\varepsilon]$



Central limit theorem

Law of large numbers: when the size of sample increases the mean of sample converges to the actual population mean.

Find confidence interval

confidence = 95%

$\sigma = 5.4$ $\quad$ $Var = 5.4^2$

$n = 50$ $\quad$ $\bar{x} = 79$

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Confidence interval $= [\bar{x} - \varepsilon, \bar{x} + \varepsilon]$

Confidence level $= 95$

$\therefore A_L = \dfrac{0.95+1}{2}$

$= 0.975$

$z = 1.96$

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$= \dfrac{5.4}{\sqrt{50}}$

$= 0.763$

Confidence interval $= [\bar{x} - \varepsilon, \bar{x} + \varepsilon]$

$\varepsilon = z_\alpha \sigma_{\bar{x}}$

z_α = z corresponding to confidence level

confidence level $= \dfrac{\text{confidence}+1}{2}$

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Confidence level	Z-score
90%	1.645
95%	1.96
98%	2.33
99%	2.575

Markov's inequality

Let X be a non-negative random variable like height, distance, etc

$$\boxed{P(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon}}$$

Chebychev's inequality

$$P(|X| > \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$$

$$P(|X| > \varepsilon) = P(X^2 \geq \varepsilon^2) \leq \frac{E(X^2)}{\varepsilon^2}$$

$$P(|X - E(X)| \geq \varepsilon) = P((X - E(X))^2 \geq \varepsilon^2 \leq \frac{E(X - E(X))^2}{\varepsilon^2}$$

$$= \frac{V(X)}{\varepsilon^2}$$

$$\therefore \boxed{P(|x-\mu| \geq \varepsilon) \leq \frac{Var(x)}{\varepsilon^2}}$$

Hoeffdings Inequality

$$\boxed{P(\bar{X}_n - \mu \geq \varepsilon) \leq e^{-\frac{2n\varepsilon^2}{b-a}}}$$

For a bernoulli experiment

$$\varepsilon = \frac{1}{\sqrt{n}} \sqrt{\frac{1}{2} \ln\left(\frac{2}{\alpha}\right)}$$

Where $\alpha = 1 - \text{Confidence}$

Eg: if confidence iterval is 95%

$$\alpha = 1 - 0.95$$
$$= \underline{\underline{0.05}}$$

Risk

Risk minimization

$$\underset{\theta}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^{n} \left[y - f(x_i ; \theta) \right]^2$$

Random number generation

Linear Congruential generator (LCG)

$$X_{n+1} = (aX_n + c) \bmod M$$

Seed: seed is the start value. The first number will be calculated based on the seed

$0 \leq seed < M$

Eg: suppose $a = 3$ $\quad c = 5$
$X_n = 7$ $\quad M = 10$

$$X_{n+1} = (3 \times 7 + 5) \% 10$$
$$= 26 \% 10$$
$$= 6$$

$$X_{n+2} = (3 \times 6 + 5) \% 10$$

$= 3$

and so on.....

Limitation of LCG: It starts to generate the same series of number after certain number of iteration. The number of iteration it takes before starting to loop is called the period

Hull-Dobell theorem

- C and m are co prime
- (a-1) is divisible by all prime factors of m
- (a-1) is divisible by 4 if m is divisible by 4

Accept reject Sampling

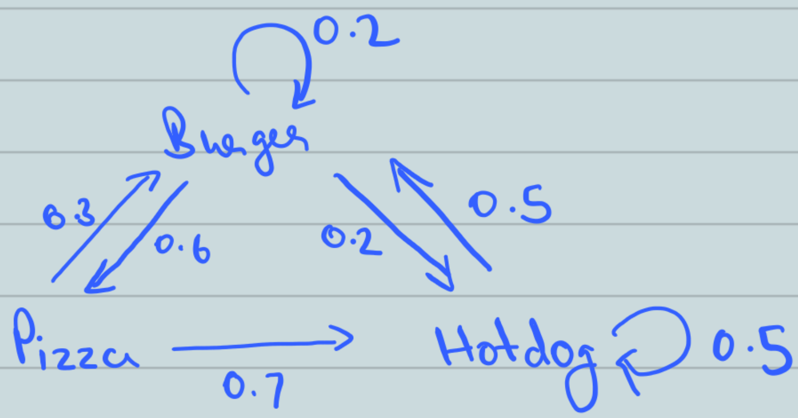
Suppose consider the function shown below

This is a function which is difficult to sample from. So we introduce a new function called proposal function

Consider $g(x)$ which is a uniform distribution. We will sample from $g(x)$ such that $f(x)$ is satisfied

Markov chain

Eg:



$$P(X_{n+1} = x / X_n = x_n)$$

ie Probability of the next state will only depend on the current state and not the states before

$$P(X_n = \text{hotdog} / X_{n-1} = \text{pizza}) = 0.7$$

We can express these probabilities using an adjacency matrix

$$A = \begin{array}{c} \\ \text{Pizza} \\ \text{Hotdog} \\ \text{Burger} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Pizza} & \text{Hotdog} & \text{Burger} \end{array} \\ \begin{bmatrix} 0 & 0.7 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} \end{array}$$

We represent a vector π to point out current state.

Eg: if current state is pizza the
$\pi = [1 \quad 0 \quad 0]$

If we calculate πA

$$= [1 \quad 0 \quad 0] \begin{bmatrix} 0 & 0.7 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

$$= [0 \quad 0.7 \quad 0.3]$$

which give the future probabilities of other states given this state is pizza

Then you can use the result of πA to calculate probabilities for the next state.

After a few iterations you will reach a state given by

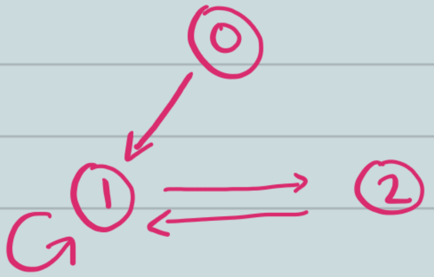
$$\boxed{\pi A = \pi} \quad - (1)$$

↳ Stationary state

$$\boxed{Ax = \lambda x} \quad - (2)$$

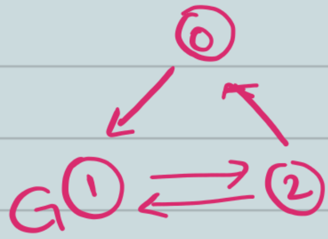
↳ Eigen vector equation

We can corelate (1) and (2) and state that (1) is represent left eigen vector



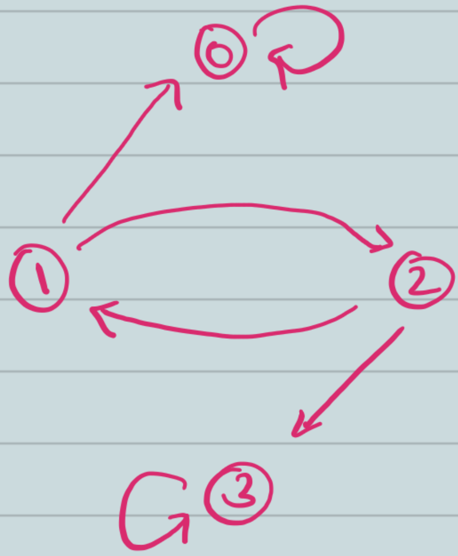
This is a reduceable markov chain. Once you move from 0 to 1 then you cannot come back to state 0 again. Also it can be reduced into 2 chains

If you add a connection between state (2) and (0) then this becomes irreducible

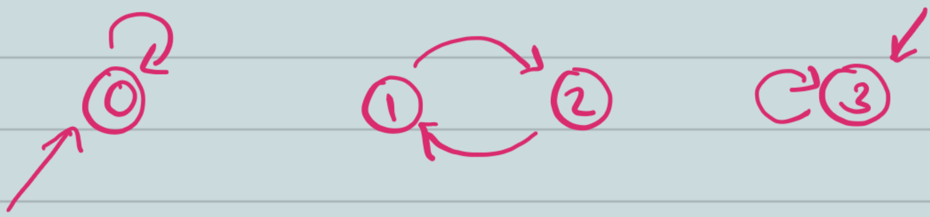


In this one all states are reachable from any given state.

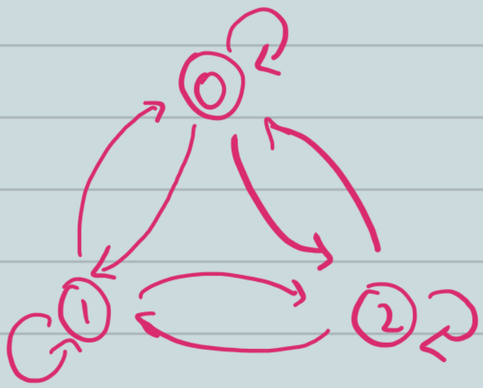
Gamblers ruin graph



this can be reduced to 3 smaller chains



Higher order transition



$$A = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} \end{array}$$

We need to find $P_{02}(1)$ which represents the probability to reach 2 from 0 in 1 step

$$P_{02}(1) = A_{02}$$

ie $$\boxed{P_{ij}(n) = A_{ij}}$$

We can generalize using the below formula

$$\boxed{P_{ij}(n) = A^n_{ij}}$$

$\lim_{n \to \infty} A^n$ will converge to stationary state

A^n will converge only if the markov chain is

* irreducible
* aperiodic

Markov chain stationary state

$$\boxed{\text{Solve for } \pi P = \pi}$$

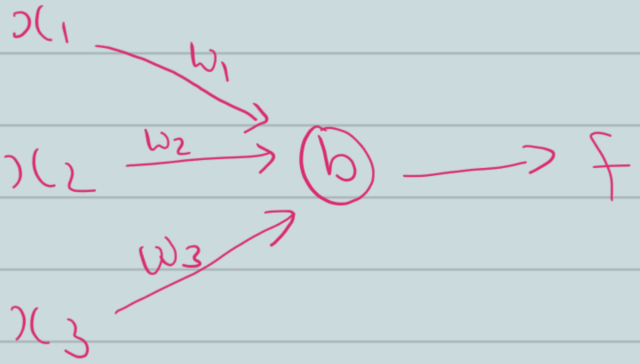
Markov chain reversibility

Prove that $\forall\ i, j$

$$\pi_i P_{ij} = \pi_j P_{ji}$$

Pattern Recognition

Perceptron: takes in multiple output and produces a binary output



if t is the threshold then we can redefine this as the below function

$$f = \begin{cases} +1 & w.x > t \\ -1 & w.x \leq t \end{cases}$$

$$\text{Precision} = P(y=1 \,/\, g(x)=1) = \frac{TP}{TP+FP}$$

$$\text{Recall} = P(g(x)=1 \,/\, y=1) = \frac{TP}{TP+FN}$$