**RESUMOS MPEI**

**Vídeo 1 (Introdução à probabilidade)**

**Experiência aleatória é constituída por:**

Espaço de amostragem - S

Conjunto de todos os resultados possíveis

Resultados têm de ser mutuamente exclusivos e não divisíveis

Discreto se for contável

Continuo se não for contável

Conjunto de acontecimentos (ou eventos) - A

É sempre um subconjunto do espaço de amostragem (S)

Lei de probabilidade

Noção frequencista -> por testagem e repetição

Axiomas de probabilidade

Aximoa1 – probabilidade são não-negativas

Axioma 2 – normalização (S tem prob 1)

Axioma 3a – Se A e B forem mutuamente exclusivos P(AUB) = P(A) + P(B)

**Probabilidades não contáveis**

Subtração de áreas ou de subtração de distâncias

**Independência**

2 acontecimentos são independentes se P(A e B) = P(A)P(B)

Dois acontecimentos que sejam mutuamente exclusivos e tenham probabilidade não nula então não podem ser independentes

0 = P(A e B) = P(A)P(B) implicaria que um deles tenha probabilidade nula

**Vídeo 2 (Probabilidade Condicional e Introdução às variáveis aleatórias)**

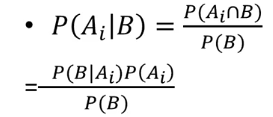
P(A e B) = P(A|B) x P(B)

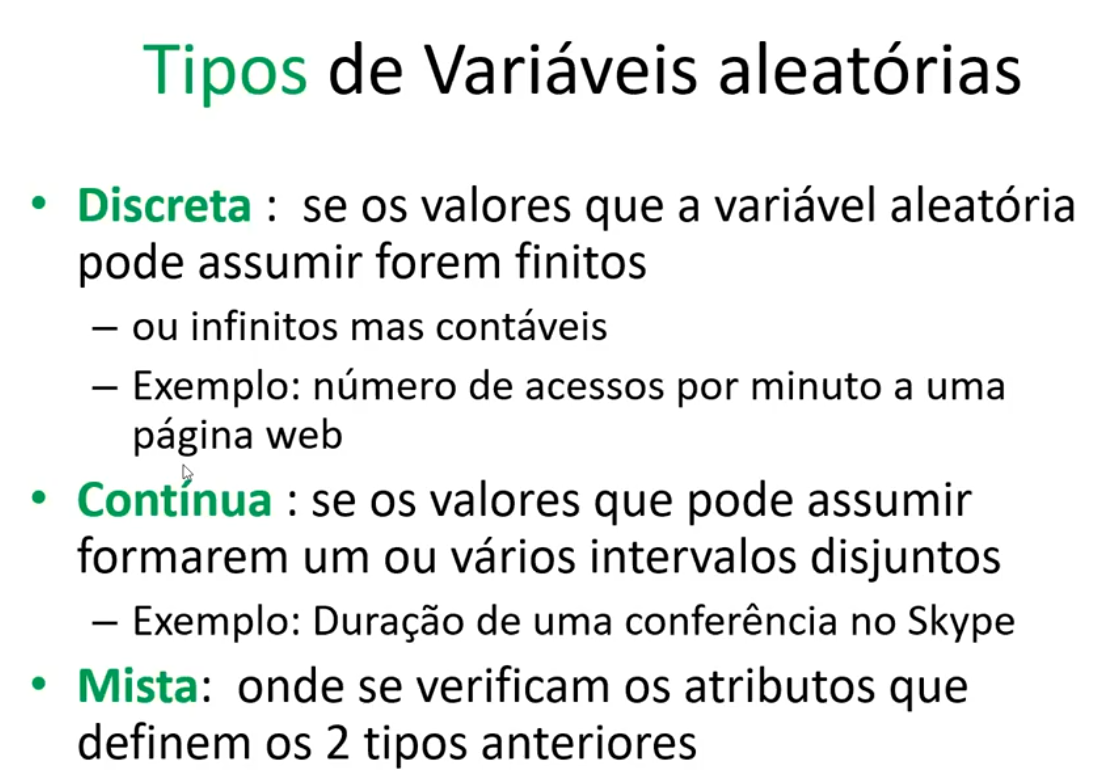
=P(A e B) x P(B) = P(A e B)

P(B)

**Regra da cadeia**

P(ABC) = P(A) P(B|A) P(C|AB)

**Regra de Bayes**

****

**Função distribuição acumulada é dada por:**

Fx(X) = Prob(X <= x)

Isto significa que para um dado de 6 faces a probabilidade de sair dois também engloba sair 1

No Matlab usa-se cumsum

**Vídeo 3 (Variáveis Aleatórias)**

Média = sum(fX.\*X)

Variância = sum(fX.\*(X.^2)) - media^2

Desvio Padrão = sqrt(variancia)

Quanto menor a variância mais perto estão os valores da média

Valor esperado = média

No caso discreto o valor esperado é dado pelo somatório de XIp(xi)

No caso continuo é dado pelo integral de -inf a +inf de xfx(x) dx

var(X+c)=var(X)

var(c X) = c2 var(X)

**Problema #1**

1 Rei, 1 Dama, 1 Valete – 1e cada

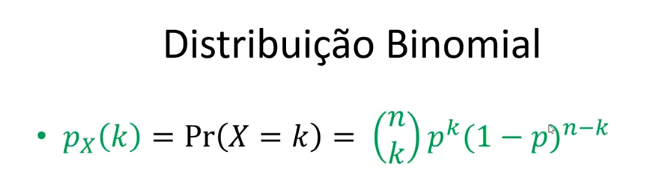
1 As – (-1e)

1 Joker – (-2e)

Como é honesto a probabilidade de sair qualquer carta é 1/5

Valor esperado = 13 – 1 – 2 = 0

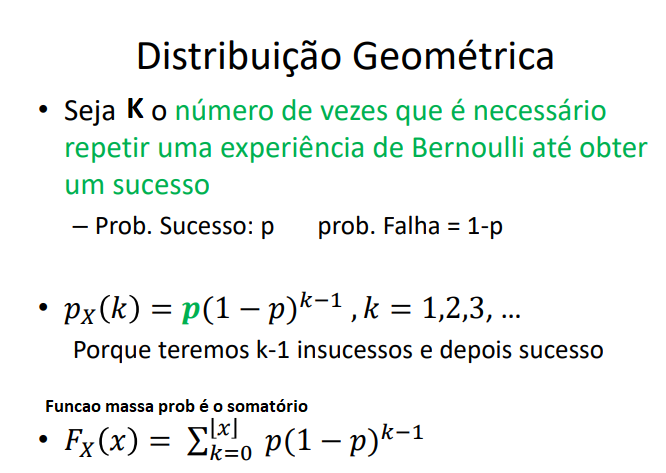
**Vídeo 4 (Distribuição Binomial)**

****

N representa o número de experiências

K representa o número de sucessos

P representa a probabilidade



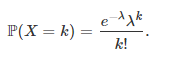
**Geométrica**

Valor médio = 1/p

Variância = (1-p)/p2

**Distribuição de Poisson**

Usada para os mesmos casos da distribuição binomial mas para N a tender para infinito e P a tender para 0

****Lambda é n\*p

Valor esperado = Lambda

Var = Lambda

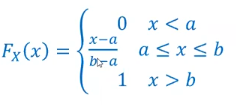
**Rule of thumb**

Se n>20 e np<=7 a aproximação de Poisson é suficientemente próxima para ser usada em vez da Binomial

**Distribuições continuas**

Distribuição Uniforme

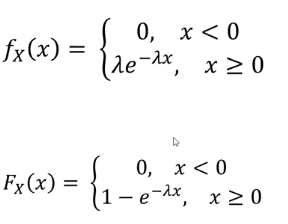
Função densidade probabilidade Função distribuição probabilidade

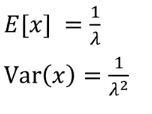


P(x1<= X <=x2) = x2 – x1 / b - a

U(a, b) significa que estamos perante uma distribuição uniforme com média A e variância B

Distribuição Exponencial (relacionada com a Poisson)





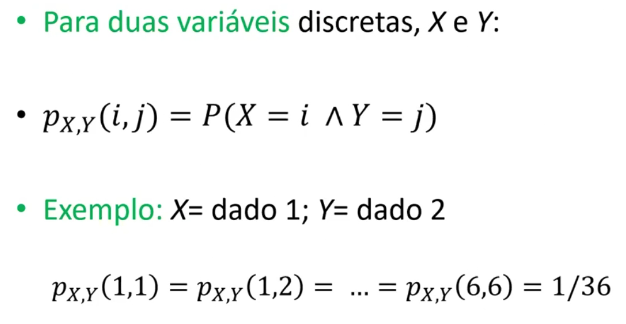
Distribuição Normal

**GAUSSIANA E NORMALIZACAO? QUE ISSO? NÃO ENTENDO NEM QUERO ENTENDER :)**

**Vídeo 5 (Variáveis aleatórias multidimensionais)**

Para caracterizar variáveis n-dimensionais, tal como nas unidimensionais em que utilizamos as funções: massa de probabilidade (para variáveis discretas), função distribuição (para ambos os tipos de variáveis) e densidade de probabilidade (para variáveis contínuas), passamos a ter as funções de distribuição conjunta:

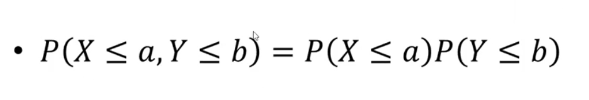
* De massa probabilidade conjunta
* De distribuição cumulativa conjunta
* De densidade de probabilidade conjunta

Função massa de probabilidade conjunta:

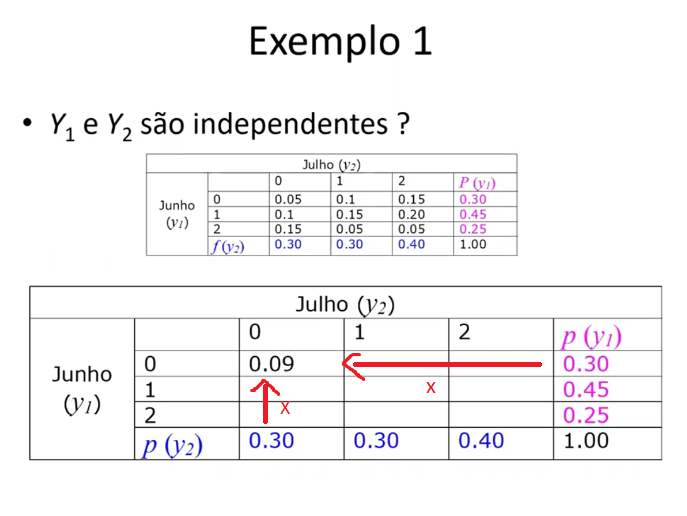
Função de distribuição acumulada:

FX,Y(x,y) = P(X <= x e Y <= y)

Se as variáveis forem independentes:



FX,Y(a,b) = FX(a)\*FY(b) (equivalente à função anterior)



**Em variáveis aleatórias**

Momento central de segunda ordem para uma das variaveis (de j ou k, sendo o outro igual a 0) = variância = E[(X – E[X])j (Y - E[Y])k]

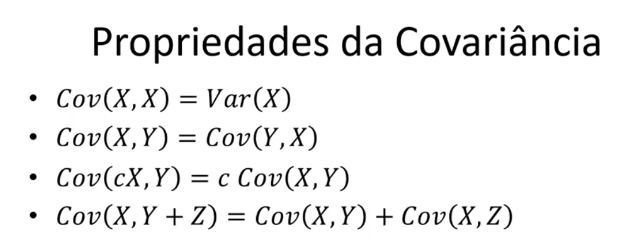
Se j=0 e k=2 obtemos a variância de Y. Se j=2 e k=0 obtemos a variância de X

**Correlação:** E[X1Y1], se a correlação é 0 diz-se que as variáveis são ortogonais

Sendo X e Y independentes E[XY] = E[X] x E[Y]

**Covariância**: é o primeiro momento das duas variáveis ou seja,

Cov(X, Y) = E[(X – E[X])1 (Y – E[Y])1] = E[XY] – E[X] x E[Y]



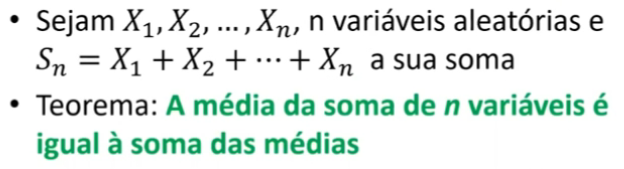
O coeficiente de correlação (ρ) é dado por:

Cov(X, Y) Se o ρxy = 0, então as variáveis são descorrelacionadas

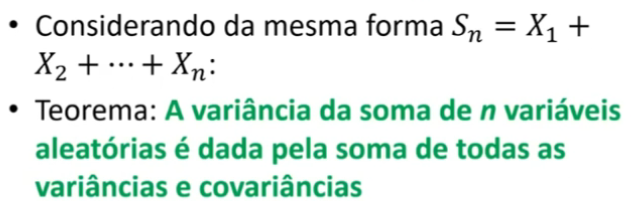
DesvioPadrão(X) \* DesvioPadrão(Y)

**Vídeo 6 (Soma de variáveis aleatórias)**

**Soma de variáveis aleatórias**

**Média:**

E[X+Y] = E[X] + E[Y]

**Variância:**

Var(X+Y) = Var(X) + Var(Y)

Se as variáveis forem independentes, a Covariância é 0, portanto a variância da soma é dada pela soma das variâncias (e não das variâncias + covariâncias, porque Cov = 0).

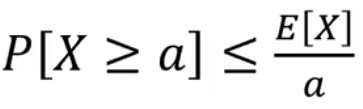
Se houver uma soma de variáveis iguais então:

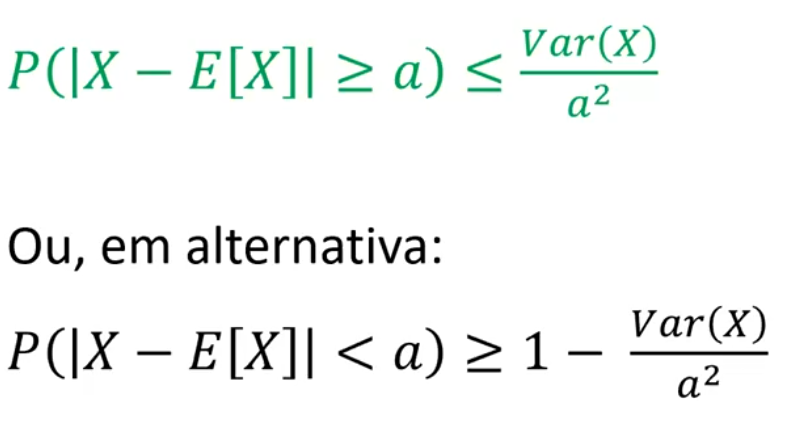
E[cXi] = c \* E[Xi]

Var(cXi) = c2 \* Var(Xi)

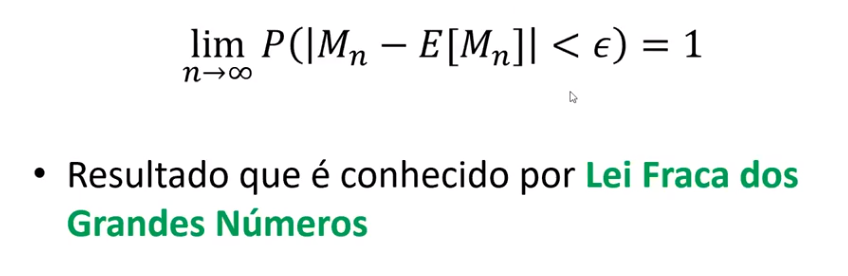
**Contadores estocásticos**

De maneira a gastar menos memória, utilizamos um contador que salta de X em X eventos, dependendo da probabilidade. Por exemplo, se a nossa probabilidade for 1/6, o contador vai saltar de 6 em 6 eventos de maneira a melhorar o tempo de execução e ocupar menos memória, abdicando, no entanto, da precisão dos resultados. O valor do contador vai representar a 2N-1 acontecimentos.

**Desigualdade de Markov**

**Desigualdade de Chebyshev**

**Vídeo 7 (Lei Fraca dos Grandes Números e Teorema do Limite Central)**



Teorema do Limite Central: Para valores grandes de *n*, temos sempre uma distribuição com a forma da Gaussiana (Normal), seja qual for a distribuição de partida (Bernoulli, Exponencial, Uniforme).

Logo a média das variáveis aleatórias, a média amostral vai ser = a das variáveis aleatórias originais.

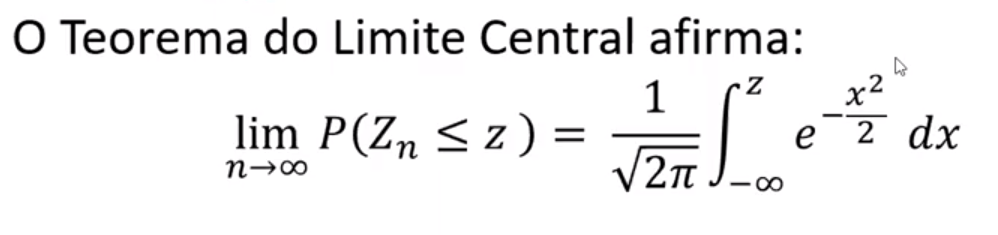
I.I.D. (Independentes Identicamente Distribuídas) --> As variáveis têm todas a mesma distribuição e a mesma função massa (se forem variáveis discretas) /densidade (se forem variáveis contínuas) de probabilidade.

S*n* -> Soma das *n* primeiras variáveis.

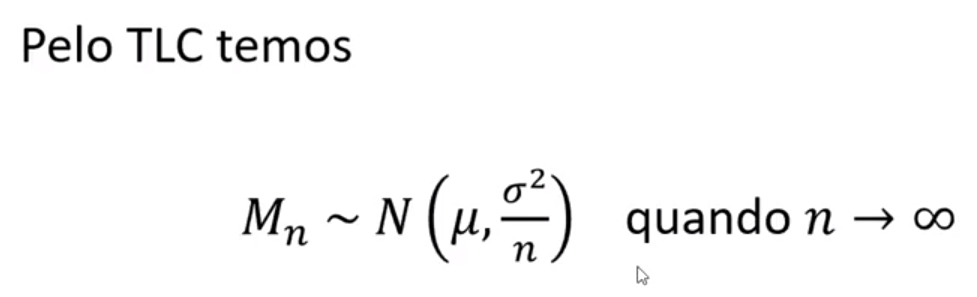
*n*μ -> Média de Sn porque E[Xi] = μ.

σ \*sqrt(*n*) -> desvio padrão.

É uma variável aleatória de média nula e variância unitária.

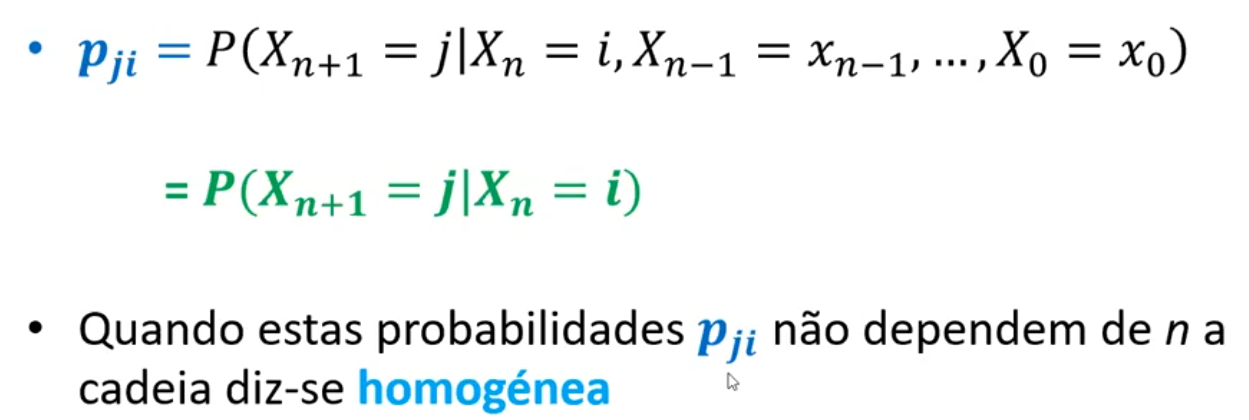


O limite da função distribuição acumulada da variável aleatória Zn será igual à função distribuição acumulada de uma variável aleatória gaussiana (normal normalizada, ou seja, média nula (= 0) e desvio padrão unitário).

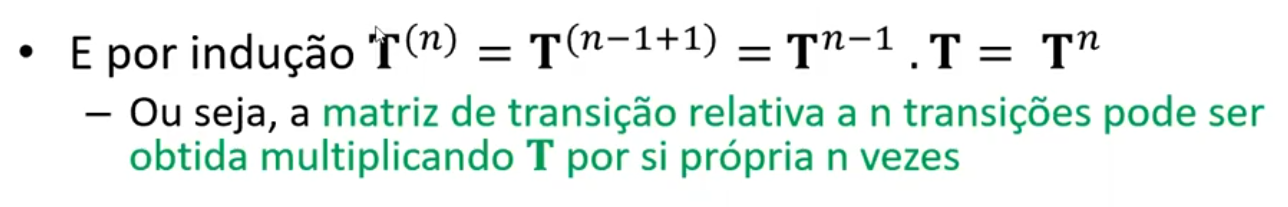


Para a média acontece o mesmo, pois consideramos esta média uma variável aleatória.

**Vídeo 8 (Cadeias de Markov)**



Diz-se que um processo de Markov não tem memória, ou seja, a probabilidade de passar para o estado seguinte depende apenas do est ado atual e em nada dos estados passados.



Uma cadeia em que todos os estados estão comunicantes diz se que é irredutível ou não redutível ou irreducible.

Uma classe é um conjunto de estados que comunicam entre sim.

Um estado recorrente é um estado que depois de sair dele pode voltar para ele mesmo, se não for possível é transiente.

**Vídeo 9 (Cadeias de Markov)**

Definição: processo de Markov é um processo estocástico em que a probabilidade de o sistema estar num estado específico num determinado período de observação depende apenas do seu estado no período de observação imediatamente precedente.

Uma matriz é **regular** se alguma potência da matriz tem todos os valores diferentes de 0

Forma canónica tem os estados absorventes no fim da matriz (últimas colunas e últimas linhas)

Um estado é **periódico** se apenas se pode regressar a ele após um número fixo de transições superior a 1 (ou múltiplos desse número).

Em consequência, uma cadeia regular é também ergódica.

**Cadeia ergódica**: se é possível efetuar transições de qualquer estado para qualquer estado (em 1 ou mais transições).

**Na forma canónica**

A matriz Q contém as transições entre estados não absorventes

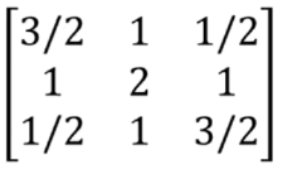
Quando multiplicamos Qn com n a tender para infinito, então Q tende para 0

**Matriz fundamental**

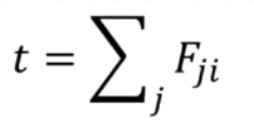
F = (I – Q)-1

A matriz Fundamental tem valores em cada entrada e esses valores representam quantas vezes esse estado vai ser visitado (em média) antes de ser absorvido.

As somas dos valores da coluna dão o número de passos (em média) até atingir um estado absorvido



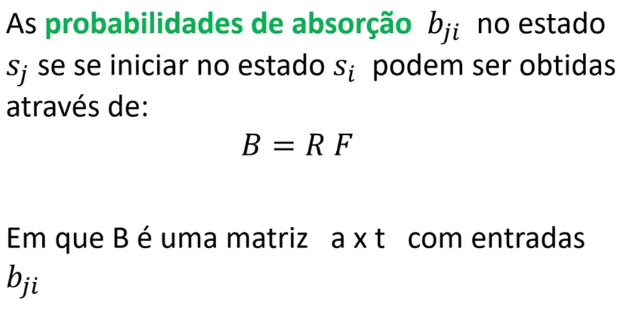
Se começarmos no estado (1,1) vamos passar em média 1,5x por esse estado antes da absorção.

Quando iniciamos o percurso no estado 1 são precisos 3 passos em média até atingir o estado absorvente

Em que j é a linha e i a coluna.

**Vídeo 10 (Cadeias de Markov)**

Probabilidades de absorção



a -> estados absorventes

t -> estados transientes

R -> Matriz por baixo da matriz Q (dos estados não absorventes para os estados não absorventes) da matriz T (de transição)

Cada valor desta matriz corresponde à probabilidade de começares num estado transiente e ires para um estado absorvente.

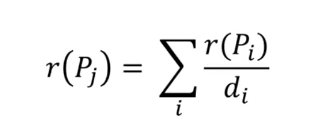
**Vídeo 11 (Pagerank)**

Num grafo, em PageRank, os nodos/vértices são as páginas web e as ligações/arcos são os hyperlinks.

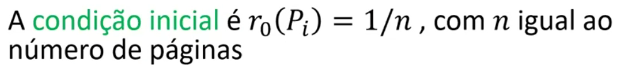
A ideia de criação de PageRank foi usar a estrutura das ligações para saber quais as páginas “importantes”.

PageRank é um algoritmo que atribui um número real a cada página da Web, o qual quanto maior mais importante é a página.

Esse número é atribuído através da seguinte expressão:



Ou seja, o PageRank de uma página é o somatório do PageRank das páginas que apontam para ela a dividir pelo número de páginas para os quais a página de que saiu, apontam.



Converge para o PageRank limite, tal como nas cadeias de Markov.

Grafo fortemente ligado: é possivel ir de qualquer página para qualquer página.

Sabemos que a dsitribuição dos pageranks atingirá um estado estacionário, em que *r* = H*r*.

Nas condições de que: o grafo é fortemente ligado e não existem becos sem saída (***dead ends*** = páginas que não têm links de saída).

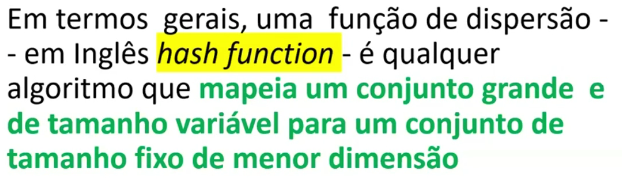
**Spider Web** = conjunto de estados que se apontam entre eles próprios, não tendo maneira de sair, como se fosse um *dead end* na forma de várias páginas.

**Vídeo 12 (PageRank)**

No caso de termos Dead Ends e/ou Spider Traps, a solução é criar ***random teleports*** que consistem em probabilidades de 1/*n*.

Função matLab para resolver dead Ends -> A (isnan(A)) = 1/n

**Vídeo 15 (Hash functions)**

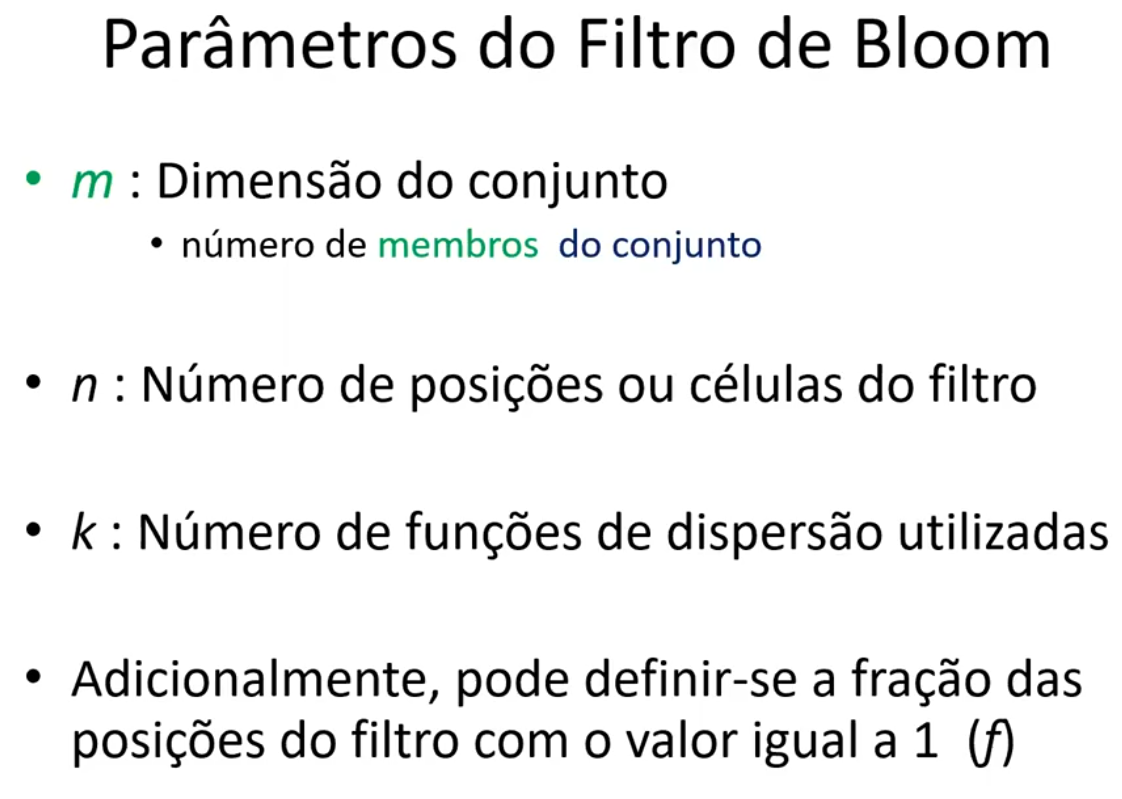


**Vídeo 16 (Filtros Bloom)**

Os filtros de Bloom usam funções de dispersão para calcular um vetor (o filtro) que é representativo do conjunto. Ou seja o valor X vai gerar k valores que vão ser preenchidos no filtro de Bloom e mudar o valor que está nesse espaço para 1 para dizer que está ocupado.

No bloom filter não há falsos NEGATIVOS

Quanto maior o tamanho do bloom filter melhor porque há menos colisões.



**Vídeo 17 (Filtros Bloom)**

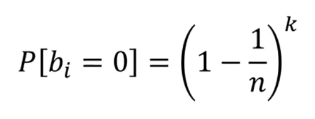
Como obter *n*? Num shei :3 :$ \*-\*

Probabilidade de um falso positivo:

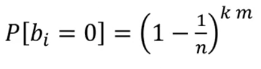
1º - probabilidade de um determinado bit estar a 1.

P[bi=1] = 1/*n* (assumindo que são equiparáveis)

2º - probabilidade de um determinado bit estar a 0, após k funções de dispersão

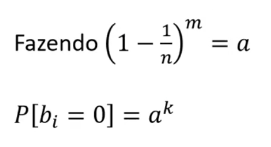


3º-probabilidade após inserir *m* elementos.



4º - probabilidade de um alvo ser atingido por pelo menos um dardo (caso exemplo)

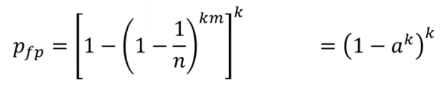
Mudança de variável:



Como nós queremos o caso em que bi = 1:



5º - Probabilidade de um falso positivo (ou seja, ter *k* bits iguais a 1 para um elemento não pertencente, após inserir *m* elementos):



Que pode ser aproximada por:



Para obter um kótimo:

