Distribuições 2

Distribuicao Binomial: 2

Poisson 3

Gaussian/Normal Distribution 4

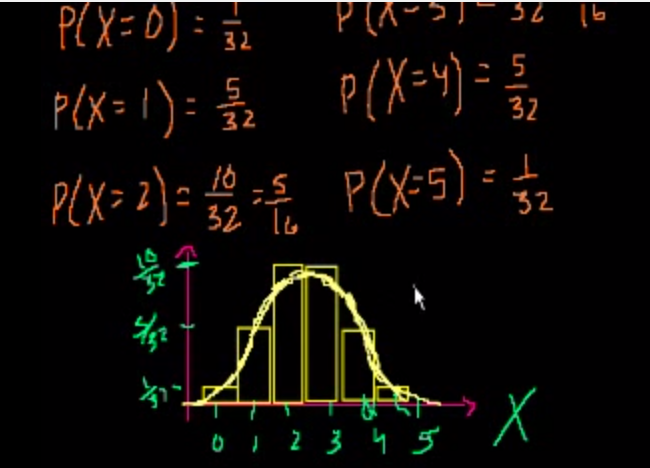
GAUSSIAN MIX 6

## Distribuições

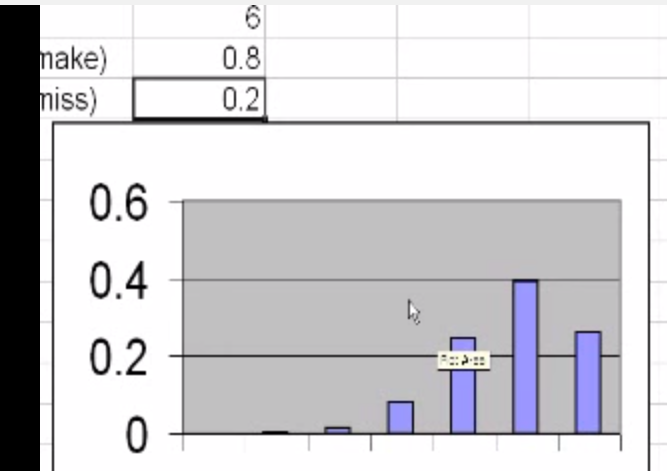
### Distribuicao Binomial:

*A Normal é um tipo de distribuição binomial... eu conto o numero de uma coisa e vejo como ela se comporta ... posso* ***assumir*** *q ela se comporta numa distribuição gaussiana/normal.*

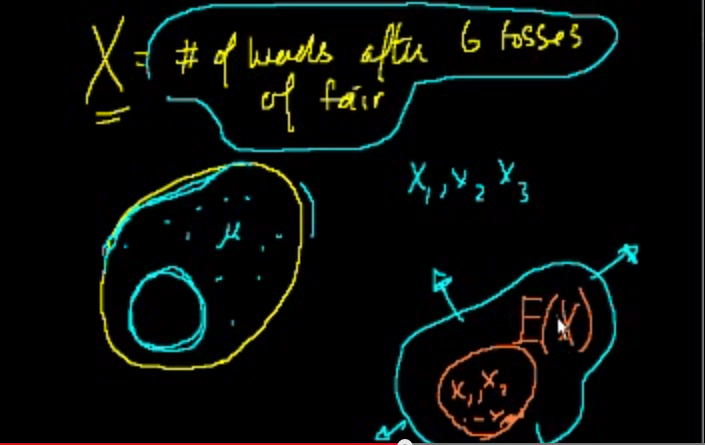
*Enfim, é binomial qdo eu conto o numero de vezes q uma coisa acontece (ou não acontece)*



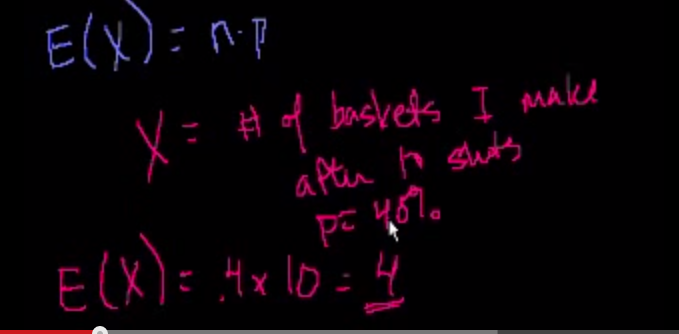
*Exemplo de acertar uma cesta de basquete*



*A esperança de uma variável randômica (expected value) é como se fosse a média de toda a população (de todos os experimentos ... q poderiam ser finitos) (ver desenho abaixo)*



*O valor esperado numa distribuição binomial é fácil:*



*Essa eh a formula da probabilidade binomial:*



*P= a prob. De acertar um tiro. Logo a prob. De acertar K tiros tem a ver com a combinação de ordens x as probabilidades de acertar e não acertar*

### Poisson

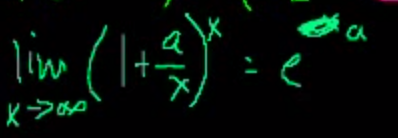
*Se eu quero o numero de carros em 1 hora.... posso faze-lo por minuto como uma binomial.*

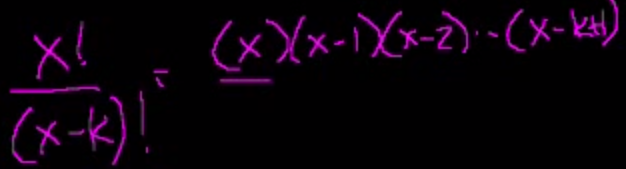
*E se 2 ou mais carros passarem no minuto? Eu ainda teria apenas 1 caso no binomial.*

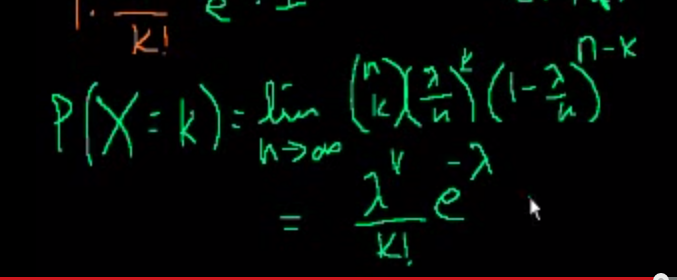
*Logo eu teria q baixar pra segundo. Ainda assim poderia ter 5 carros passando no mesmo segundo.... e assim sucessivamente eu teria q ir baixando a granularidade do tempo.*

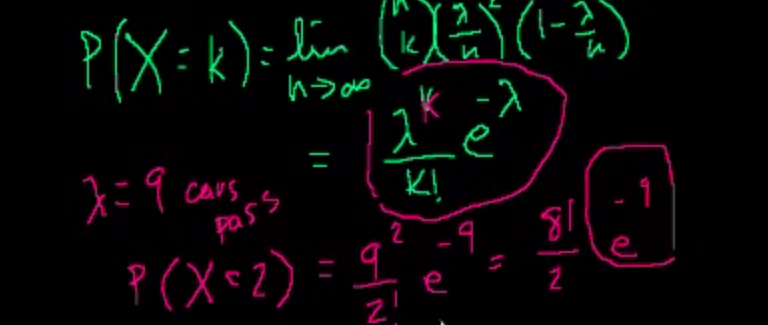
*Essa é a idéia do Poison.... ele não deixa de ser uma distribuição BINOMIAL*

*Antes, Math:*





 *essa eh a formula da distribuição de Poisson.... onde o lambda é o valor esperado E(Y)=lambda=n.p*



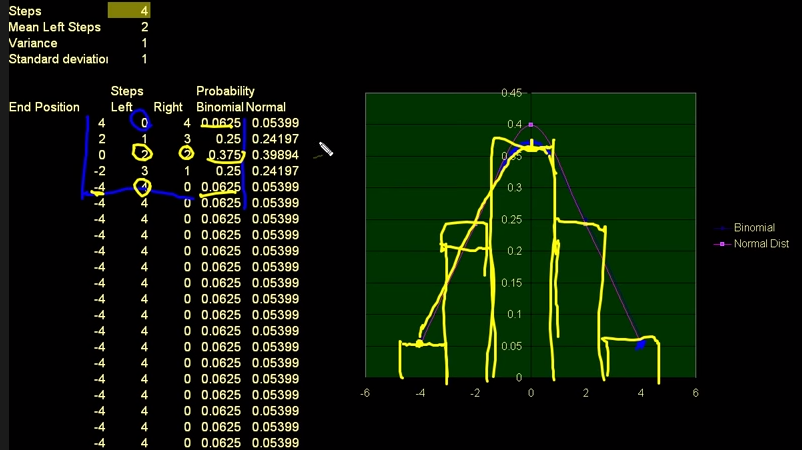
*(é uma dist. Binomial... com a diferença de que ele granulariza ao infinito...)*

*Pra fechar... exatamente o q foi dito... q o valor esperado é para ser a média da populacao*

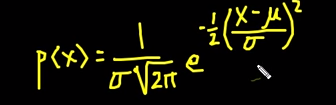
### Gaussian/Normal Distribution

*(na verdade é uma aproximação da binomial se executada muitas vezes... por isso o e^(...) ... a binomial converge pra Normal.)*

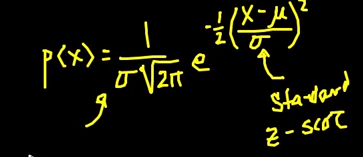
*Por isso é tão boa essa distribuição ... porque vc se aproxima de uma qtidade muito grande de vezes.... e a Gaussiana é contínua... já a normal é discreta*



*Essa é a formula da distribuição normal (probability density function to the normal distribution):*



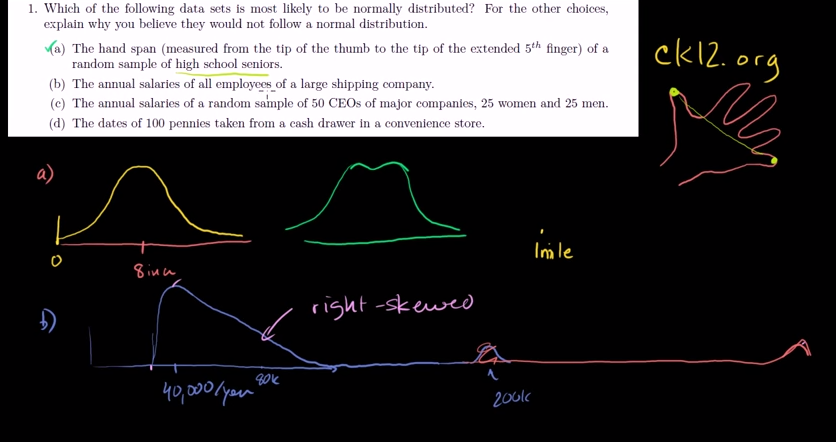
*O Z-score mede qtos desvios padrões estamos longe da media*



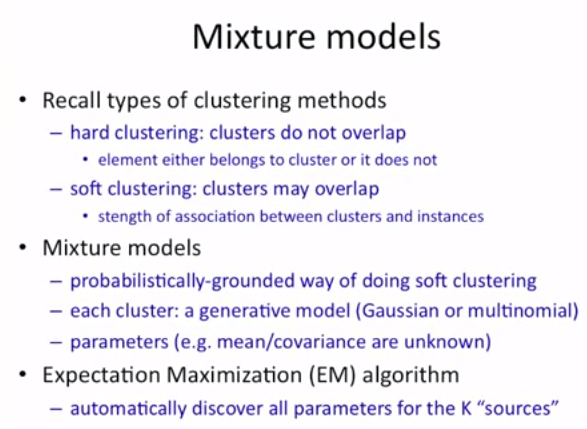
*\*\*\* vc precisa saber a variância e a média para calcular o x*

*\*\*bimodal quer dizer q tem 2 modas, 2 picos*

*\*\*right-skewed --- é uma curva com um rabo maior pra direita (não é normal)*



## GAUSSIAN MIX



*\*1 – mixture models prob grounded to soft clustering*

*\*2- cada cluster é uma distribuicao de prob. (gaussiana ou multinomial)*

*\*3 e o q eu quero descobrir com o algoritmo são os parâmetros: mean/covariance ...*

*Expectation Maximization* ou EM, pertece a uma classe de técnicas estatísticas para a estimativa de parâmetros para modelos estatísticos quando existem variáveis latentes (ou escondidas). Os parâmetros encontrados são estimativas por máxima verossimilhança ou *maximum likelihood estimates (MLE)*.

Much of their popularity can be attributed to the existence of efficient and robust procedures

for learning parameters from observations. Often, however, the only data available for

training a probabilistic model are incomplete (or Missing values can occur, for example, in medical diagnosis, where patient histories generally include results from a limited battery of tests --- aqui no meu caso eh um numero limitado de escolhas...)

**The expectation maximization algorithm enables parameter estimation in probabilistic models with incomplete data.**

Our goal is to estimate θ= (θA,θB) que é a probabilidade de ser pro-custo e contra-custo. We refer to z as hidden variables or latent factors. Parameter estimation in this new setting is known as the incomplete data case. Isso porque, no meu caso, eu nao sei se a escolha veio de um grupo pro-Custo ou anti-custo.... e quero justamente decidir qual a chance (ou melhor) as duas distribuicoes que separam as 2 opcoes.... eu vou saber a prob. de um novo custo, por causa da media e variancia dessas normais

*#the unobserved latent factors z are the assignments of genes to clusters,*

#Gaussian distributions representing the expression patterns for each cluster.

The expectation maximization algorithm is a natural generalization of maximum likeli-

hood estimation to the incomplete data case.

what cluster does a particular gene most likely belong to?

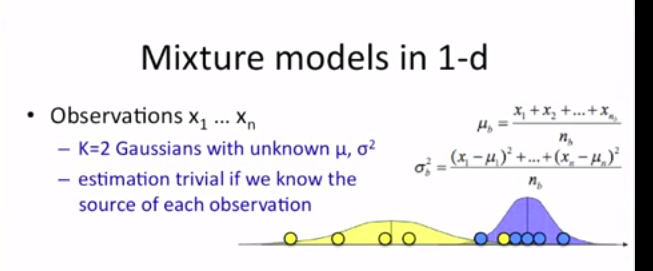
 Finalmente, a verossimilhança de um conjunto de parâmetros é, basicamente, a probabilidade desses parâmetros gerarem os resultados obtidos. O MLE é um método de inferência dos parâmetros com verossimilhança (ou probabilidade) máxima. O EM é uma generalização, portanto, do MLE em cenários que existem variáveis latentes.

EM is an algorithm for maximizing a likelihood function when some of the variables in your model are unobserved (i.e. when you have latent variables).

Como mencionado anteriormente, um dos primeiros requisitos é o de que exista variável latente no modelo a ser estudado, seja ela ausente na observação, ou seja ela propositalmente escondida. Além disso é necessário saber a função de distribuição de probabilidade da qual se deseja descobrir os parâmetros. Para alguns cenários essa função não é tão evidente e pode demandar certo esforço para ser corretamente definida.

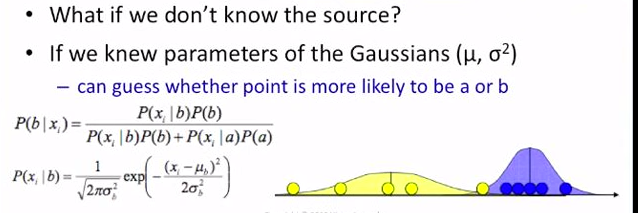
BEM! Eu posso assumir q é gaussiana.

No exemplo abaixo: k=2 (fonte amarela e azul) e qdo eu seia origem da fonte, fica fácil descobrir os meus parâmetros (q são média e variância). Eu sei de onde vem cada ponto... de qual distribuição.

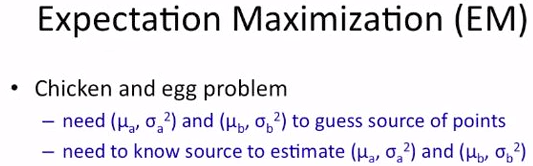


CASO 2

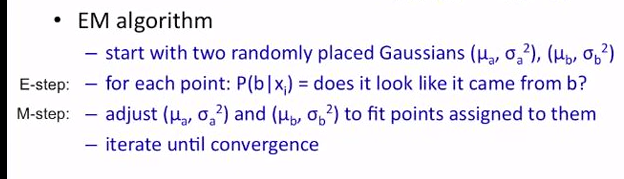
(eu não sei a fonte, mas sei os parâmetros.... posso inferior os pontos. Basta calcular a probabilidade de Bayes para saber qual é a prob de um ponto pertencer a uma distribuição ou outra. Sendo que a prob. de x ser de uma ou outra distribuição é dada pela própria formula da Normal.



There's a chicken-and-egg problem in that to solve for your model parameters you need to know the distribution of your unobserved data; but the distribution of your unobserved data is a function of your model parameters. E vc nao sabe de nenhum dos dois….

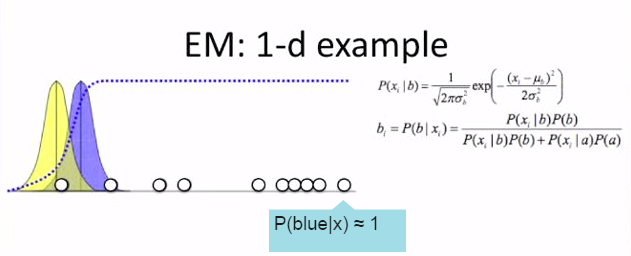


E-M tries to get around this by iteratively guessing a distribution for the unobserved data,



E-step: For each datapoint that has missing values, use your model equation to solve for the distribution of the missing data given your current guess of the model parameters and given the observed data (note that you are solving for a distribution for each missing value, not for the expected value). Now that we have a distribution for each missing value, we can calculate the expectation of the likelihood function with respect to the unobserved variables. If our guess for the model parameter was correct, this expected likelihood will be the actual likelihood of our observed data; if the parameters were not correct, it will just be a lower bound.

M-step: Now that we've got an expected likelihood function with no unobserved variables in it, maximize the function as you would in the fully observed case, to get a new estimate of your model parameters.

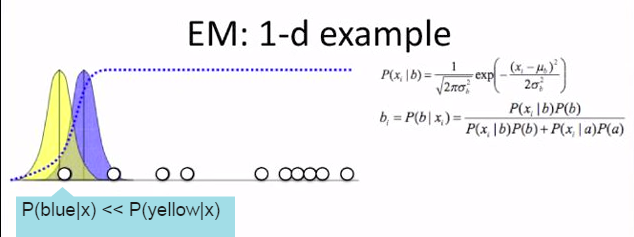


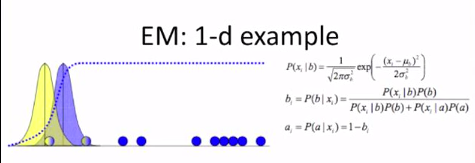
P(x|b) – a probabilidade de x ser azul .... basta jogar na formula da distribuição gaussiana.

(posso assumir os priors como iguais por enquanto)

A probabilidade daquele ponto ser azul é baixa... mas eh ainda menor. Entao, ele ficara com prob=1 para azul porque é mais provável do q amarelo.

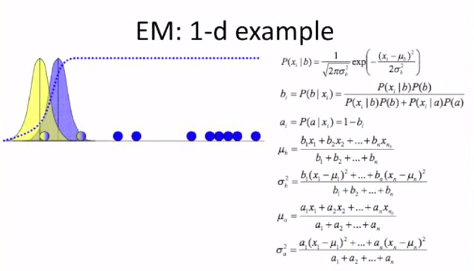
P(b|x) – a probabilidade de ser azul, levando em conta que x poderia ser azul ou amarelo (1-azul)?

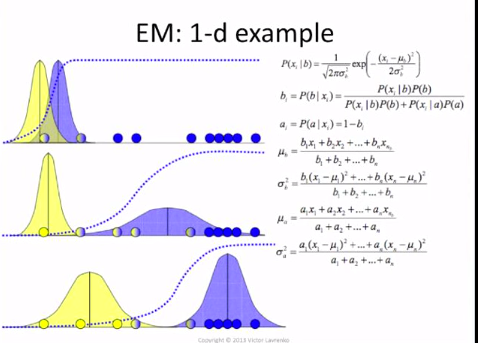




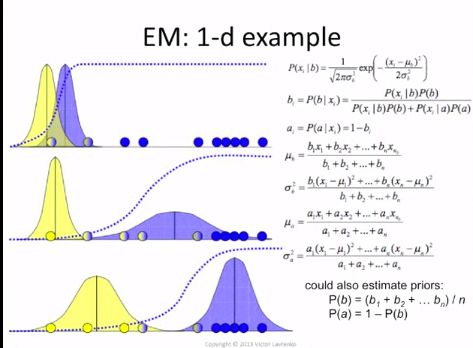
A diferença do K-means aqui é q no EM cada ponto tem uma prob pra cada centroide, já no Kmeans, ou eh num ou noutro.

Entao recalculo tudo:

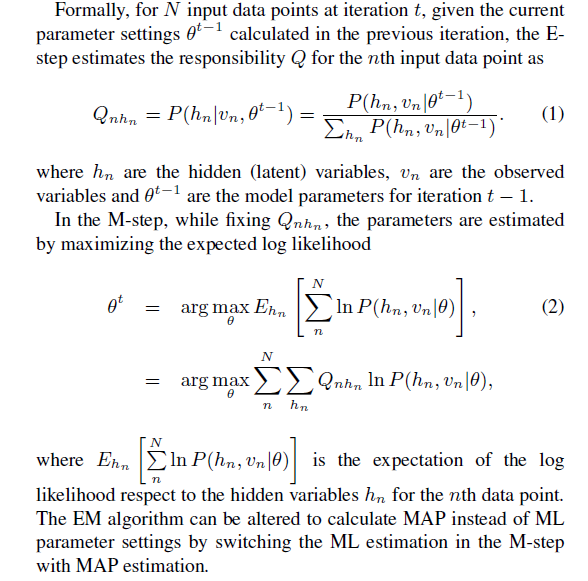




Opaaaa ainda posso estimar os priors baseado na informação de perfil, não?



REF A



mixture model

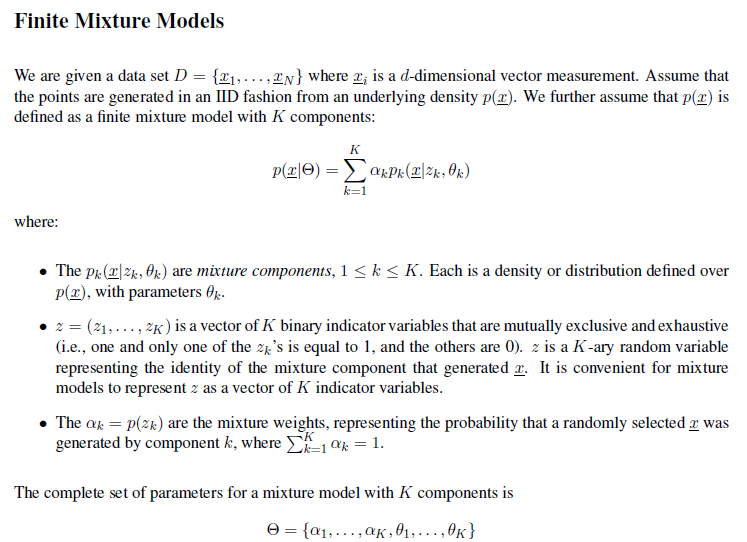
for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs. Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population.

"mixture models" are used to make statistical inferences about the properties of the sub-populations given only observations on the pooled population, without sub-population identity information.

1. D={xt, t=1,2,...,N}

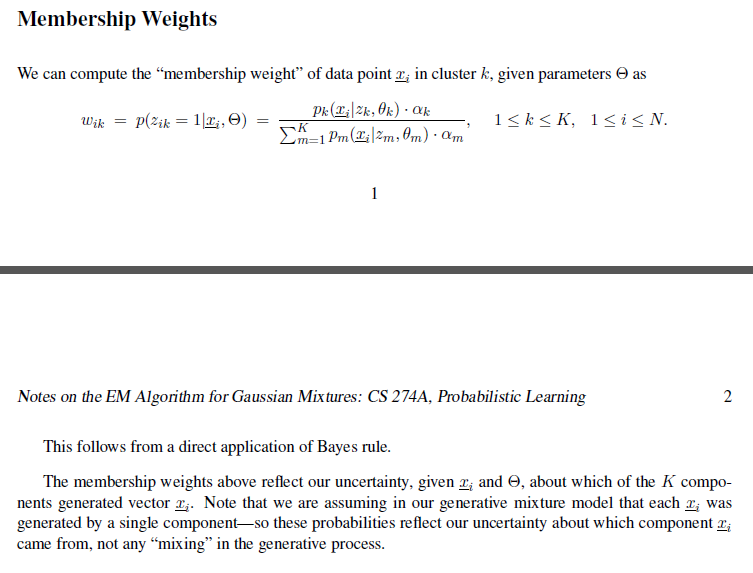
2. Subs x\_1 por x vetor \_t

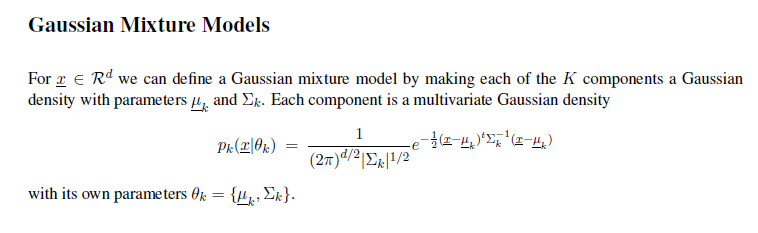
-colocar ref melhor botar o Bilmes e Barber



In [probability theory](http://en.wikipedia.org/wiki/Probability_theory) and [statistics](http://en.wikipedia.org/wiki/Statistics), a [sequence](http://en.wikipedia.org/wiki/Sequence) or other collection of [random variables](http://en.wikipedia.org/wiki/Random_variable) is **independent and identically distributed** (**i.i.d.**) if each random variable has the same [probability distribution](http://en.wikipedia.org/wiki/Probability_distribution) as the others and all are mutually [independent](http://en.wikipedia.org/wiki/Independence_(probability_theory)).[[1]](http://en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables#cite_note-1)

The [abbreviation](http://en.wikipedia.org/wiki/Abbreviation) *i.i.d.* is particularly common in [statistics](http://en.wikipedia.org/wiki/Statistics) (often as *iid*, sometimes written *IID*),





(nao usei )

unobserved (i.e. when you have latent variables).

Cluster:

the unobserved latent factors z are the assignments of genes to clusters, .... We refer to z as hidden variables or latent factors

Gaussian distributions representing the expression patterns for each cluster.... what cluster does a particular gene most likely belong to?

is a probabilistically-grounded way of doing soft clustering....

each cluster is a generative model: Gaussian or multinomial (for example)

There's a chicken-and-egg problem in that to solve for your model parameters you need to know the distribution of your unobserved data; but the distribution of your unobserved data is a function of your model parameters. E vc nao sabe de nenhum dos dois….

need (mi e desvio)A e (mi e desvio)B to guess the source of points

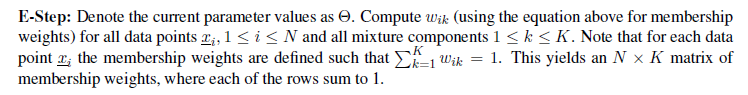
and need the source of points to estimate (mi e desvio)A e (mi e desvio)B

E-M tries to get around this by iteratively guessing a distribution for the unobserved data,

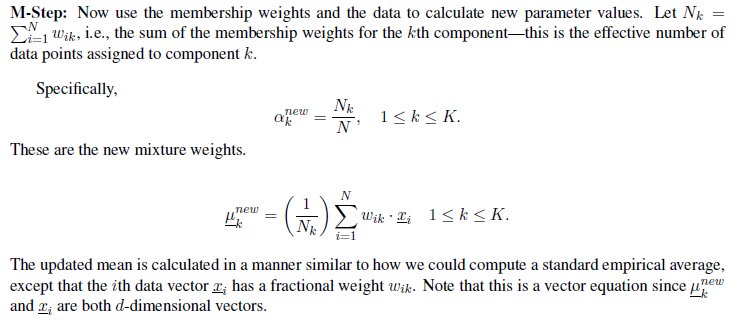
alternate between estimating the unknowns

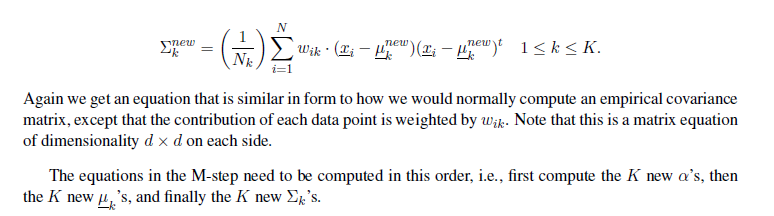
parameters and the hidden variables (z).

Ref A



Ref B





Ref C

