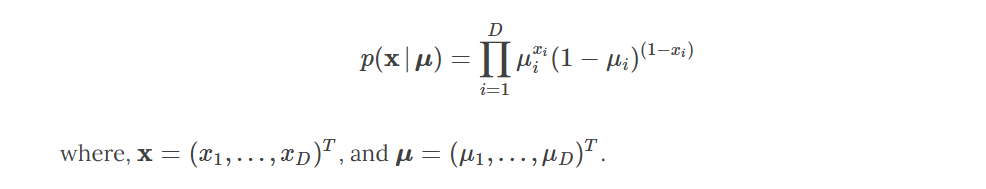
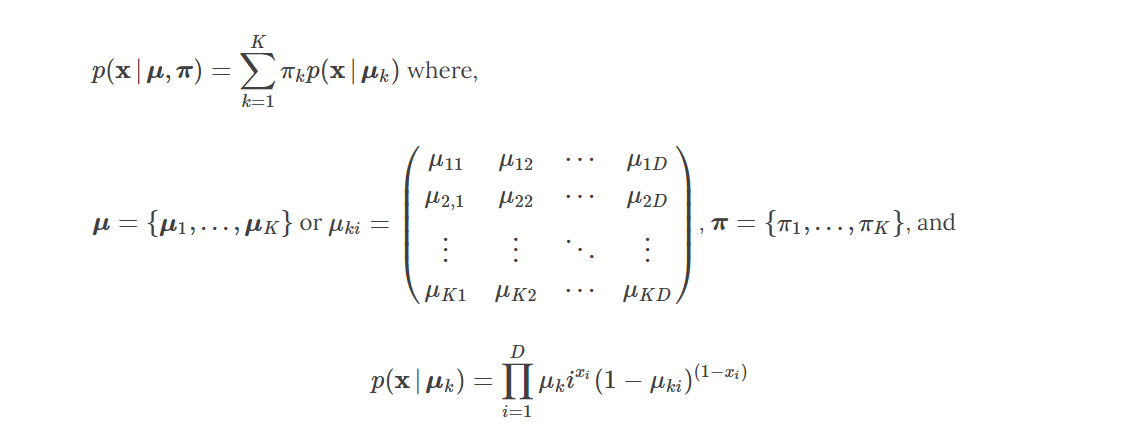
**EM on Binomial Mixture Models**

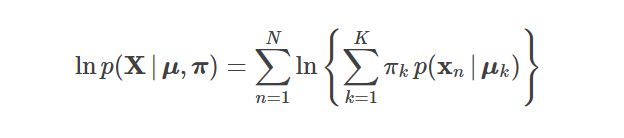
First, consider a single multivariate random variable x with Bernoulli distribution of D independent binary variables xi∈{0,1}, where i=1,...,D, each of which is in turn a univariate Bernoulli distribution with parameter μi,



Now consider a finite mixture of K multivariate Bernoulli distributions given by,

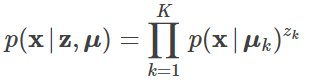


Given a data set of X={x1,...,xN}, with each observation represented as a mixture of K Bernoulli distributions, then the log likelihood function is:

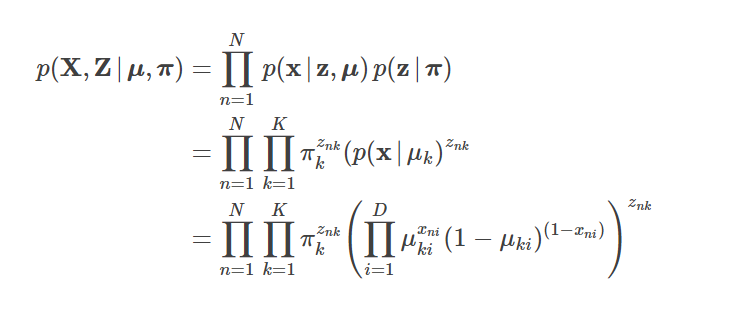


The appearance of the summation inside the logarithm in the above function means maximixum likelihood solution no longer has a closed form solution.

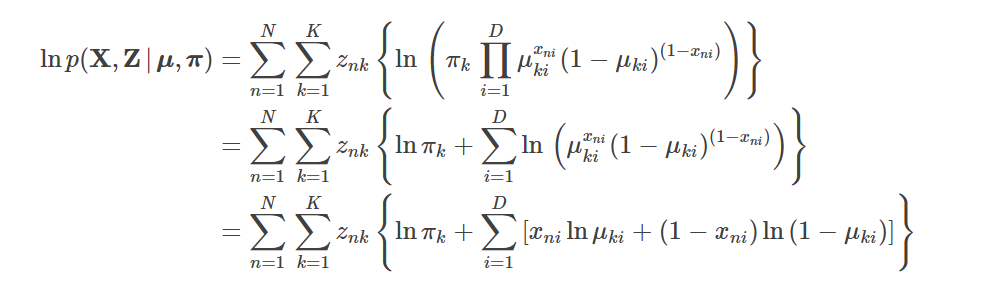
To maximize the log likelihood function using Expection-Maximization approach, consider an explicit latent variable z associated with each observation x, where z=(z1,...,zK)T is a binary K-dimensional variable having a single component equal to 1, and all others set to 0. The latent variable z can be considered as a membership indicator for each observation. The marginal distribution of z is specified in terms of the mixing coefficients πk, such that p(zk=1)=πk. We can write the conditional distribution of x, given the latent variable as



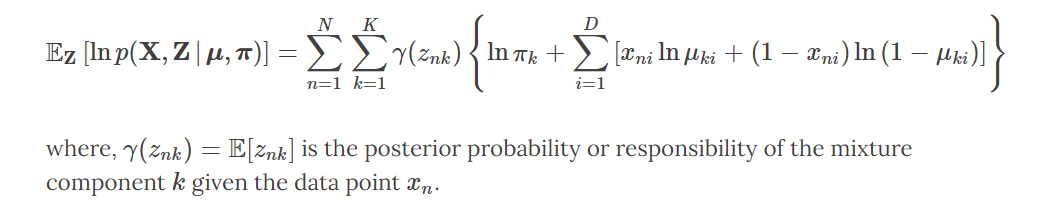
Now formulate the probability of the complete-data (observed x and latent z) using Bayes’ theorem, p(x,z)=p(x|z)p(z). For the complete-data, the probability is



The corresponding complete-data log likelihood functions is:



Notice that the above log likelihood function can be considered as a linear combination of znk. Since the expectation of a sum is the sum of the expectations, we can write the expectation of the complete-data log likelihood functions with respect to the posterior distribution of the latent variable as:

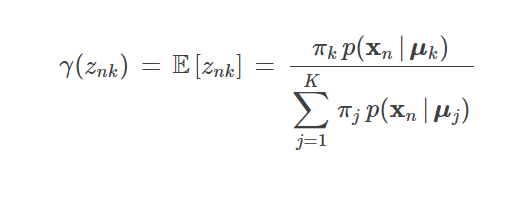


where, γ(znk)=E[znk] is the posterior probability or responsibility of the mixture component k given the data point xn.

**E-M Algorithm**

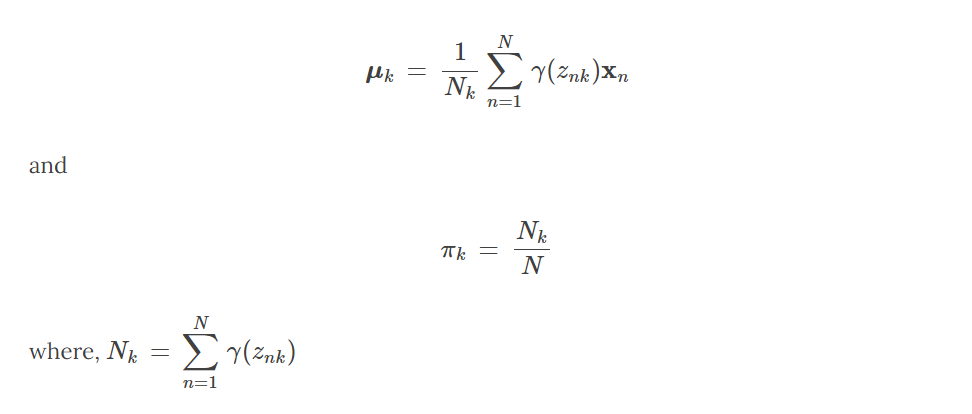
1. **E step**

Calculation of the responsibilities make the E step of the E-M algorithm.



1. **M step**

Maximizing the expectation of the complete-data log likelihood with respect to μk and πk yields the M step of the E-M algorithm:



**Link to the collab file**

[**https://colab.research.google.com/drive/1VREHnpzVA79KZAuoKwZADk7JkAW6R4zu?usp=sharing**](https://colab.research.google.com/drive/1VREHnpzVA79KZAuoKwZADk7JkAW6R4zu?usp=sharing)