With the finite mixture model-based approach to density estimation and clustering, the density of is modeled as a mixture of a number of component densities in some unknown proportions. The model can be mathematically described as a FMM on the data points, where it is unknown which mixture, or subpopulation, each data points belongs to. In other words, a finite mixture model is a statistical model that assumes the presence of unobserved latent groups within an overall population. Each latent group can be fit with its own regression model, which may have a linear or generalized linear response function.

It is assumed in FMM that there are a total of mixture components, such that a data point belongs to component (or subpopulation) j with probability In other words, the different subpopulations correspond to a partition such that each data point in the population belongs to one and only one subpopulation. Each data point in subpopulation has a capture probability , assumed to be constant over time. The set of capture probabilities is given by .

Mixture models are flexible tools that allow modelling the associated structure of a set of variables (their joint density) using a finite mixture of simpler densities. Mixture models have a single latent variable (called indicator variable*)* that points to the mixture component. Mixture components can be modeled using a prior distribution for mixing proportions that selects a reasonable subset of components to explain any finite training set (Neal R. 1991). Formally, the mixture model is defined in the following way:

(1.12)

Here*,*  is the probability that the example is in cluster *;*  is the probability of if is in cluster . The generative process of the mixture model consists of cluster sampling and sampling data examples from the cluster distribution. Mixture models infer mixtures of distributions for each component separately. In most cases, the goal to compute posterior distributions of parameters or latent variables requires multidimensional integration.Deriving posterior means or simply identifying regions of high posterior density value pose highly complex computational challenges. Standard Monte Carlo methods are available for simulating from a posterior distribution associated with a mixture, allowing implicit integration over the entire parameter space. By the [law of large numbers](https://en.wikipedia.org/wiki/Law_of_large_numbers), integrals described by the [expected value](https://en.wikipedia.org/wiki/Expected_value) of some random variable can be approximated by taking the [empirical mean](https://en.wikipedia.org/wiki/Sample_mean_and_sample_covariance) (the sample mean) of independent samples of the variable. When the [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) of the variable is parametrized, [Markov chain Monte Carlo](https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo) (MCMC) sampler may be used. As to draw independent samples from a joint posterior distribution is computationally intractable, a Markov chain is constructed, i.e., MCMC generates samples from the posterior distribution by constructing a reversible Markov chain. It is expected that in a very long run samples will take values that look like draws from the target distribution (at equilibrium). By the [ergodic theorem](https://en.wikipedia.org/wiki/Ergodic_theorem), the stationary distribution is approximated by the [empirical measures](https://en.wikipedia.org/wiki/Empirical_measure) of the random states of the MCMC sampler (Stack Exchange 2018). Various algorithms are used for posterior sampling (i.e., sampling to generate the posterior distribution) – for high-dimensional Gaussian distributions, Gibbs sampler may be an option in order to approximate the hidden variable distributions. These determine the next steps of the Markov chain.

Clustering procedures based on finite mixture models are being increasingly preferred to[*heuristic methods*](https://www.sciencedirect.com/topics/mathematics/heuristic-method) due to their sound mathematical basis and to the [interpretability](https://www.sciencedirect.com/topics/mathematics/interpretability) of their results. Mixture model-based procedures provide a probabilistic clustering that allows for overlapping clusters corresponding to the components of the mixture model. The uncertainties that the observations belong to the clusters are provided in terms of the fitted values for their posterior probabilities of component membership of the mixture. As each component in a finite mixture model corresponds to a cluster, the problem of choosing an appropriate [clustering method](https://www.sciencedirect.com/topics/mathematics/clustering-method) can be recast as statistical model choice. It also allows the important question of how many clusters are there in the data to be approached through an assessment of how many components are needed in the mixture model. These questions of model choice can be considered in terms of the likelihood function.