Example of Graphical Model – Mixture Model and Hidden Markov Model –

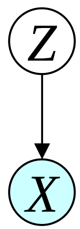
Kenji Fukumizu
The Institute of Statistical Mathematics

Computational Methodology in Statistical Inference II

Finite Mixture Model

Mixture Model

- Graphical model of finite mixture model



Hidden variable (unobservable)

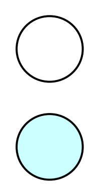
Observable variable

The distribution of X depends on the hidden variable Z.

Z: discrete variable taking value in $\{1,2,\ldots,K\}$

X: either discrete or continuous

Convention in this course:



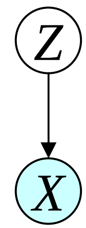
blank circle – hidden variable

colored circle – observable

Mixture Model

- Probability density of finite mixture model
  - Joint probability

$$p(X,Z) = p(Z)\,p(X \mid Z)$$



$\{1,2,\ldots,K\}$

  - Marginal of X

$$\begin{aligned} p(X) &= \sum_{Z} p(Z)\,p(X \mid Z) \\ &= \sum_{k=1}^{K} p(Z=k)\,p(X \mid Z=k) \\ &= \sum_{k=1}^{K} \pi_k\, p_k(X) \end{aligned}$$

$$\pi_k := p(Z=k)$$

$$p_k := p(X \mid Z=k)$$

General form: $$p(X) = \sum_{k=1}^{K} \pi_k\, p_k(X)$$

$\sum_{k=1}^{K} \pi_k = 1,\ \ \pi_k \geq 0,\ \ p_k(x)$: density of X

Examples of Mixture Model

- The components p_k are often taken from a popular family of probabilities.

- Gaussian mixture model

$$p(x) = \sum_{j=1}^{K} \pi_j \phi(x \mid \mu_j, \Sigma_j)$$

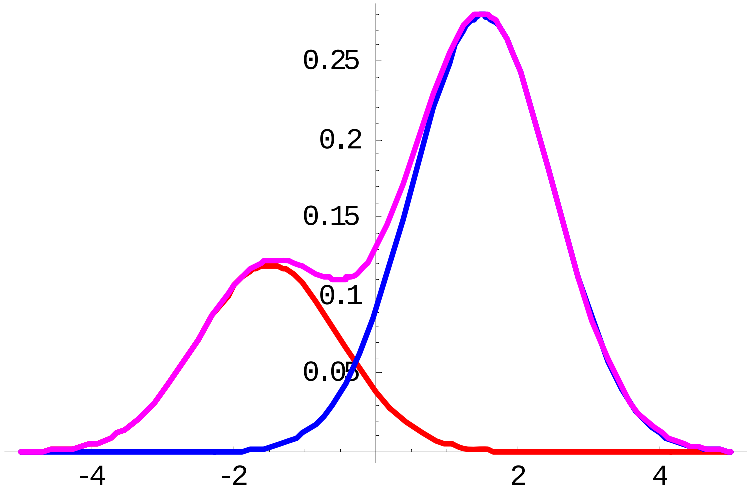
where $\phi(x \mid \mu, \Sigma)$: density function of normal distribution with mean μ and covariance matrix Σ

i.e.,

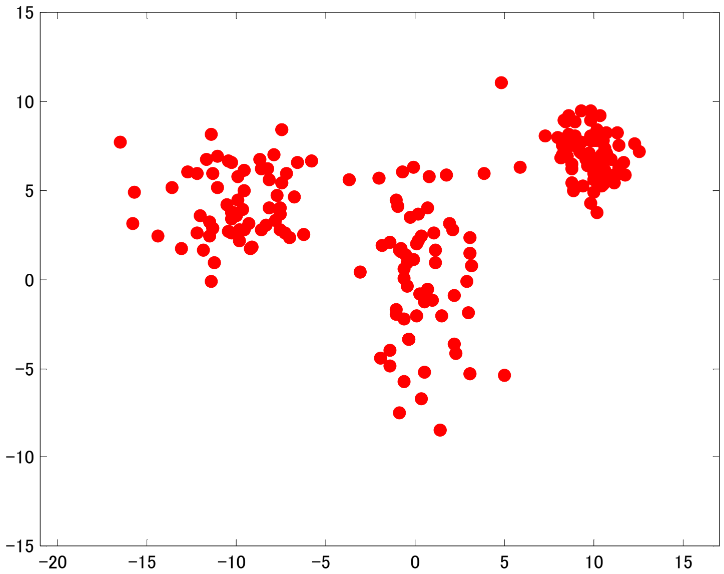
$$\phi(x \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \mid \Sigma \mid}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

- Mixture of binomials, mixture of chi-squares, etc

Gaussian Mixture Model



P.d.f . of Gaussian mixture
1 dimensional, 2 components.



I.i.d. sample from Gaussian Mixture
2 dimensional, 3 components,
200 data.

Application of Mixture Model

- Gaussian Mixture
  - Modeling of clustered data
  - Statistical foundation of analyzing clustering
  - Outlier detection, etc....

- Others
  - Mixture of binomial distributions:

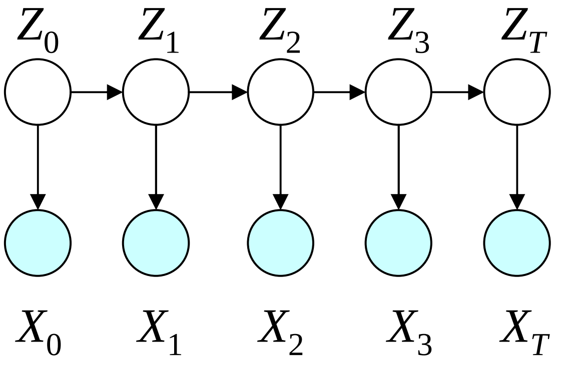
    Statistical model for linkage analysis in genetics.
    The ratio of combination may be different over different groups.

Estimation of the parameter from data will be discussed later (Part IV).

Hidden Markov Model

Hidden Markov Model (HMM)

- Graphical model of HMM



Z_i: hidden state, often discrete

X_i: observable

- Probability density of HMM

$$p(X,Z) = p(Z_0)\prod_{t=1}^{T} p(Z_t \mid Z_{t-1})p(X_t \mid Z_t)$$

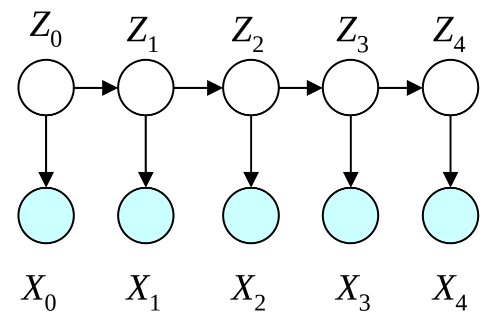
$$p(X) = \sum_{Z_0}\sum_{Z_1}\cdots\sum_{Z_T} p(Z_0)\prod_{t=1}^{T} p(Z_t \mid Z_{t-1})p(X_t \mid Z_t)$$

Hidden Markov Model

- State transition

$$p(X) = \sum_{Z_0,\ldots,Z_T} p(Z_0)\prod_{t=1}^{T} p(Z_t \mid Z_{t-1})p(X_t \mid Z_t)$$

  - The probability $p(Z_t \mid Z_{t\text{-}1})$ is the transition probabilities of K states.



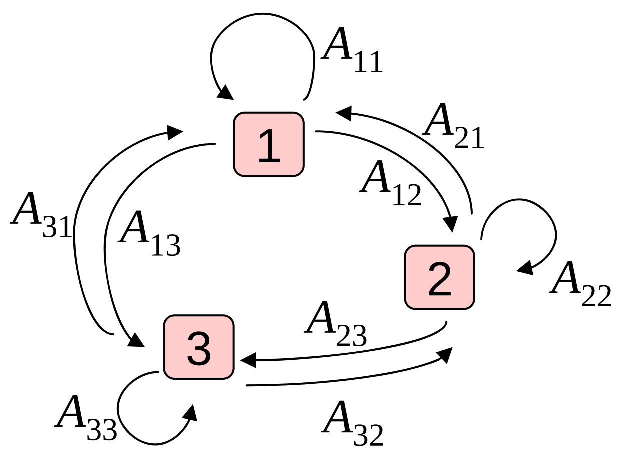
$$A_{jk}^{(t-1)} = p(Z_t = k \mid Z_{t-1} = j)$$

$$A_{jk}^{(t)} \geq 0,\ \sum_{k=1}^{K} A_{jk}^{(t)} = 1$$

They are often time-invariant:

$$A_{jk}^{(t)} = A_{jk}$$

  - Transition diagram



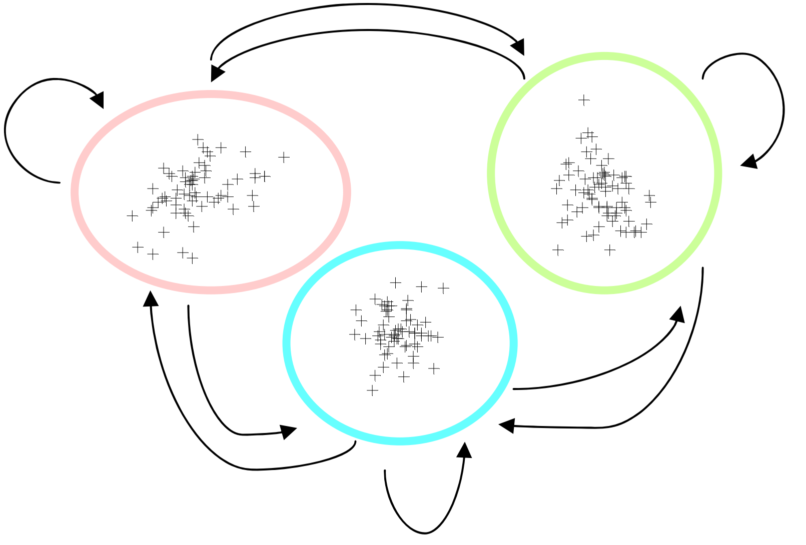
Hidden Markov Model

- Example

  Gaussian hidden Markov model

  $p(X_t \mid Z_t = j\ )$ is Gaussian

$$p(x \mid Z = j) = \phi(x \mid \mu_j, \Sigma_j)$$



- If the hidden state is generated independently, HMM is equal to a mixture model.

- If the state is continuous, the model is often called state-space model.

Applications of HMM

- Speech signal processing
  - Speech signals are often modeled by HMM.
    Speech recognition etc.

  (See, e.g., tutorial: Rabiner. *Proc. IEEE*, 77(2), 257–286, 1989.)

- Genome sequence
  - DNA: symbol sequence of {A, T, G, C}
  - Protein sequence: symbol sequence of 20 amino acids

  (See, e.g., Durbin, Eddy, Krogh, Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1999.)

- Natural language processing

Prediction, Smoothing, Filtering

- Inference with HMM
  - Prediction:

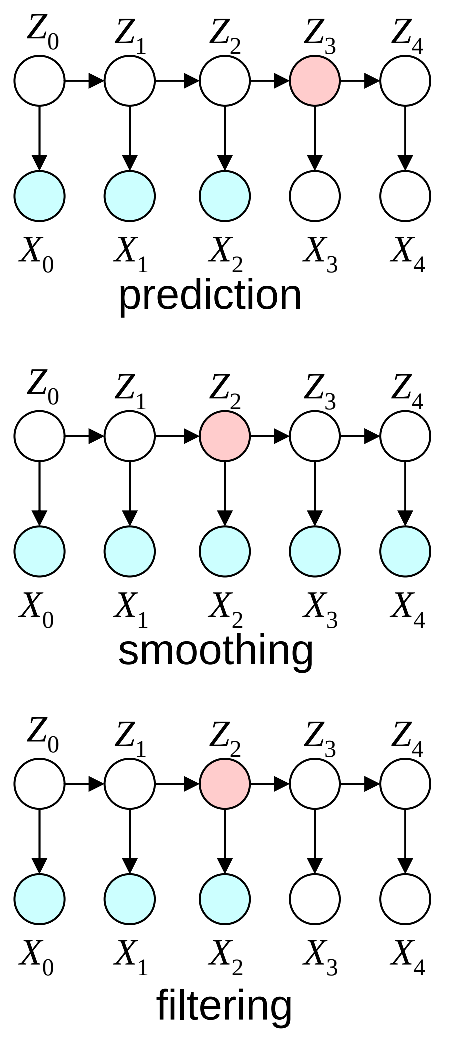
$$p(Z_t \mid X_{0,}, \ldots, X_s) \quad (s < t)$$

  - Smoothing:

$$p(Z_t \mid X_{0,}, \ldots, X_u) \quad (u > t)$$

  - Filtering:

$$p(Z_t \mid X_{0,}, \ldots, X_t)$$



Prediction, Smoothing, Filtering

- Computational difficulty

  To obtain

$$p(Z_s \mid X_0, \ldots, X_t) = \frac{p(Z_s, X_0, \ldots, X_t)}{p(X_0, \ldots, X_t)} \qquad \text{(prediction)}$$

  we need to compute

$$p(X_0, \ldots, X_t) = \sum_{Z_0} \sum_{Z_1} \cdots \sum_{Z_t} p(Z_0) \prod_{i=1}^{t} p(Z_i \mid Z_{i-1}) p(X_i \mid Z_i)$$

  Direct computation requires K^t operations – exponential on t.

- Efficient algorithms (discussed later in Part III and IV)
  - Computation of $p(X)$: forward-backward algorithm
  - Computation of most likely hidden sequence: Viterbi algorithm
  - Estimation of parameters: Baum-Welch algorithm

How to Work with Graphical Models

How to Work with Graphical Models?

- Determining structure
  - Structure given by modeling

    e.g. Mixture model, HMM
  - Structure learning → Part IV

- Parameter estimation
  - Parameter given by some knowledge
  - Parameter estimation with data such as MLE or Bayesian estimation → Part IV

- Inference
  - Computation of posterior and marginal probabilities → Part III