International Workshop on Statistical-Mechanical Informatics 2007 (IW-SMI 2007) IOP Publishing
Journal of Physics: Conference Series **95** (2008) 012023 doi:10.1088/1742-6596/95/1/012023

A quantum-statistical-mechanical extension of Gaussian mixture model

Kazuyuki Tanaka[1,†] **and Koji Tsuda**[2]

[1]Graduate School of Information Sciences, Tohoku University, 6-3-09 Aramaki-aza-aoba, Aoba-ku, Sendai 980-8579, Japan

[2]Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany

E-mail: †`kazu@smapip.is.tohoku.ac.jp`

Abstract. We propose an extension of Gaussian mixture models in the statistical-mechanical point of view. The conventional Gaussian mixture models are formulated to divide all points in given data to some kinds of classes. We introduce some quantum states constructed by superposing conventional classes in linear combinations. Our extension can provide a new algorithm in classifications of data by means of linear response formulas in the statistical mechanics.

1. Introduction

Statistical approaches are applied to data mining in computer sciences. Data in real worlds include some fluctuations and it is expected to employee some statistical methods to do systematical approaches for the treatments of such data. Many statistical approaches for inferences in data mining are based on Bayesian formulas and maximum likelihood[1]. One of fundamental approaches is a classification by means of Gaussian mixture models. In Gaussian mixture models, each data point yields from one of Gaussian distributions with averages and variances. If we hope to divide a given data to three kinds of classes, we have to prepare three kinds of Gaussian distributions.

For the classification of given data by means of Gaussian mixture models, a temperature has been introduced in order to search the maximal point of hyperparameters for the marginal likelihood by using an iterative procedure[2]. The procedure is proposed with being based on an idea of simulated annealing methods in the Markov random fields.

As one of the other physical approaches to find optimal solutions for massive probabilistic inferences, we have quantum annealing method. Quantum annealing method is formulated by replacing thermal effect by quantum effect in the simulated annealing[3, 4, 5, 6, 7]. Quantum annealing has been introduced in the probabilistic image processing as well as optimization problems[8]. It has been proved that the quantum annealing can find an optimal solution more quickly than the simulated annealing[9]. Moreover, some quantum error correcting codes are proposed and some performance limits have been given by using a gauge theory in the physics[10, 11, 12]. On the other hand, an edge state in image processing has been extended to a quantum state[13]. Thus, many physicists are interested in applications of quantum fluctuations to computer sciences.

In applying quantum effects to computer sciences, we have mainly two different viewpoints. One of them is an annealing and is applied to find an optimal solution in massive computational problems. The other is to adopt a quantum state itself as a state in computational models or probabilistic models. Our data often include a point to which it is difficult to assign one label in the statistical classification. It is interesting to introduce quantum states as label states in the classifications in both statistical and physical viewpoints.

In this paper, we extend conventional Gaussian mixture models to quantum Gaussian mixture models and give an algorithm to determine the estimates of hyperparameters in our proposed models. In section 2, we summarize conventional Gaussian mixture models and the scheme of determination of hyperparameters in the statistical framework. In section 3, we propose quantum Gaussian mixture models. The probability density functions of conventional Gaussian mixture models are expressed in terms of density matrix representations. Density matrix representations of conventional Gaussian mixture models are given by diagonal matrices. We introduce off-diagonal elements in the reformulated density matrix representation of Gaussian mixture model. In order to derive the extremum condition of the marginal likelihood in quantum Gaussian mixture model with respect to hyperparameters, we adopt linear response formulas for density matrices. In sections 4 and 5, we give some numerical experiments and concluding remarks, respectively.

2. Conventional Gaussian Mixture Model

In the present section, we explain a classification of given data by using the conventional Gaussian mixture model. The framework is based on maximum likelihood estimation and Bayesian formula in the statistics.

We introduce a set of data consisting of N real number $y_0, y_1, \cdots, y_{N-1}$. The set of data is expressed in terms of an N-dimensional column vector $\boldsymbol{y} = (y_0, y_1, \cdots, y_{N-1})^{\mathrm{t}}$. We consider to classify the given data to K kinds of classes. A label assigned to y_i is expressed by x_i. and the set of labels is expressed in terms of an N-dimensional column vector $\boldsymbol{x} = (x_0, x_1, \cdots, x_{N-1})^{\mathrm{t}}$. The variable y_i can take any real value in the interval $(-\infty, +\infty)$. An integers of $0, 1, 2, \cdots, K-1$ is assigned to the label variable x_i.

Now we consider to infer how a given data is classified to K kinds of classes by means of Bayes formula and maximum likelihood estimation. $\boldsymbol{x}$ and $\boldsymbol{y}$ can be regarded as sets of random variables for parameters and data, respectively. We assume that a set of parameters $\boldsymbol{x}$ is generated by according to the following prior probability:

$$\mathcal{P}(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{i=0}^{N-1}\sum_{k=0}^{K-1}\alpha_k\delta_{x_i,k}, \tag{1}$$

where the set $\boldsymbol{\alpha} \equiv \{\alpha_0, \alpha_1, \cdots, \alpha_{K-1}\}$ satisfies the condition $\sum_{k=0}^{K-1}\alpha_k = 1$ and $\delta_{a,b}$ is Kronecker's delta. Given data $\boldsymbol{y}$ are generated from parameters $\boldsymbol{x}$ by according to the following conditional probability:

$$\mathcal{P}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\sigma}) = \prod_{i=0}^{N-1} g_{x_i}(y_i|\boldsymbol{\mu},\boldsymbol{\sigma}), \tag{2}$$

$$g_k(y_i|\boldsymbol{\mu},\boldsymbol{\sigma}) \equiv \frac{1}{\sqrt{2\pi}\sigma_k}\exp\Big(-\frac{1}{2{\sigma_k}^2}(y_i-\mu_k)^2\Big). \tag{3}$$

In the statistics, the set $\boldsymbol{\mu} \equiv \{\mu_0, \mu_1, \cdots, \mu_{K-1}\}$, $\boldsymbol{\sigma} \equiv \{\sigma_0, \sigma_1, \cdots, \sigma_{K-1}\}$ as well as $\boldsymbol{\alpha} \equiv \{\alpha_0, \alpha_1, \cdots, \alpha_{K-1}\}$ are referred to as *hyperparameters*. Equations (1) and (2) give us the following joint probability of $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\mathcal{P}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma}) = \mathcal{P}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\sigma})\mathcal{P}(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{i=0}^{N-1} \alpha_{x_i} g_{x_i}(y_i|\boldsymbol{\sigma},\boldsymbol{\alpha}) \tag{4}$$

In the statistical point of view, the estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, $\widehat{\boldsymbol{\alpha}} \equiv \{\widehat{\alpha}_0, \widehat{\alpha}_1, \cdots, \widehat{\alpha}_{K-1}\}$, $\widehat{\boldsymbol{\mu}} \equiv \{\widehat{\mu}_0, \widehat{\mu}_1, \cdots, \widehat{\mu}_{K-1}\}$ and $\widehat{\boldsymbol{\sigma}} \equiv \{\widehat{\sigma}_0, \widehat{\sigma}_1, \cdots, \widehat{\sigma}_{K-1}\}$, can be determined from given data $\boldsymbol{y}$ so as to maximizing the marginal likelihood defined by

$$\mathcal{P}(\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma}) = \sum_{x_0=0}^{K-1}\sum_{x_2=0}^{K-1}\cdots\sum_{x_{N-1}=0}^{K-1} \mathcal{P}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma}) = \prod_{i=0}^{N-1}\sum_{x_i=0}^{K-1} \alpha_{x_i} g_{x_i}(y_i|\boldsymbol{\sigma},\boldsymbol{\alpha}) \tag{5}$$

The extremum condition of $\mathcal{P}(\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma})$ with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ can be given as

$$\widehat{\mu}_k = \frac{\sum_{i=0}^{N-1} y_i \Psi_i(k|\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}},\widehat{\boldsymbol{\alpha}})}{\sum_{i=0}^{N-1} \Psi_i(k|\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}},\widehat{\boldsymbol{\alpha}})}, \tag{6}$$

$$\widehat{\alpha}_k = \frac{1}{N}\sum_{i=0}^{N-1} \Psi_i(k|\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}},\widehat{\boldsymbol{\alpha}}), \tag{7}$$

$$\widehat{\sigma}_k^2 = \frac{\sum_{i=0}^{N-1} (y_i - \widehat{\mu}_k)^2 \Psi_i(k|\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}},\widehat{\boldsymbol{\alpha}})}{\sum_{i=0}^{N-1} \Psi_i(k|\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}},\widehat{\boldsymbol{\alpha}})}, \tag{8}$$

$$\Psi_i(k|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha}) \equiv \frac{\alpha_k g_k(y_i|\boldsymbol{\mu},\boldsymbol{\sigma})}{\sum_{k=0}^{K-1} \alpha_k g_k(y_i|\boldsymbol{\mu},\boldsymbol{\sigma})}. \tag{9}$$

By solving equations (6)-(8) in the iteration method numerically, we obtain the estimates $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\alpha}}$ for given data $\boldsymbol{y}$.

By using Bayesian formula, we derive a posterior probability of the set of parameter $\boldsymbol{x}$ for given data $\boldsymbol{y}$ as follows:

$$\mathcal{P}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha}) = \frac{\mathcal{P}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})}{\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})} = \prod_{i=0}^{N-1} \Psi_i(x_i|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha}). \tag{10}$$

The estimates of $\boldsymbol{x}$, $\widehat{\boldsymbol{x}} \equiv (\widehat{x}_1, \widehat{x}_2, \cdots, \widehat{x}_{N-1})^{\mathrm{t}}$ can be determined so as to maximizing the posterior probability $\mathcal{P}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ with respect to $\boldsymbol{x}$.

3. Quantum Gaussian Mixture Model

In this section, we extend conventional Gaussian mixture models to quantum Gaussian mixture models. Our extension is based on the construction of quantum states by superposing conventional classes in the linear combinations. We give also the determination of hyperparameters by means of the maximization of likelihood for the quantum Gaussian mixture models.

For any matrix $\boldsymbol{A}$, the exponential function $e^{\boldsymbol{A}}$ and the logarithm function $\ln(\boldsymbol{A})$ are defined by

$$e^{\boldsymbol{A}} \equiv \sum_{n=0}^{+\infty} \frac{1}{n!}\boldsymbol{A}^n, \qquad \ln\boldsymbol{A} \equiv -\sum_{n=1}^{+\infty} \frac{1}{n}(\boldsymbol{I}-\boldsymbol{A})^n. \tag{11}$$

We introduce the following two $K\times K$ real diagonal matrices F and $G(y_i)$:

$$\boldsymbol{F} \equiv -\begin{pmatrix} \ln\alpha_0 & 0 & \cdots & 0 \\ 0 & \ln\alpha_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \ln\alpha_{K-1} \end{pmatrix}, \tag{12}$$

$$\boldsymbol{G}(y_i) \equiv \begin{pmatrix} g_0(y_i) & 0 & \cdots & 0 \\ 0 & g_1(y_i) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & g_{K-1}(y_i) \end{pmatrix}. \tag{13}$$

The marginal likelihood $\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ in equation (5) is replaced by the following density matrix:

$$\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha}) = \prod_{i=0}^{N-1} \frac{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}}{\mathrm{tr}\, e^{-\boldsymbol{F}}}, \tag{14}$$

where

$$\begin{aligned} \boldsymbol{H}(y_i) &\equiv \boldsymbol{F} - \ln\boldsymbol{G}(\boldsymbol{y}_i) \\ &= -\begin{pmatrix} \ln(\alpha_0 g_0(y_i)) & 0 & \cdots & 0 \\ 0 & \ln(\alpha_1 g_1(y_i)) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \ln(\alpha_{K-1} g_{K-1}(y_i)) \end{pmatrix}. \end{aligned} \tag{15}$$

Now we extend the diagonal matrix F to any $K\times K$ real symmetric matrix as follows:

$$\boldsymbol{F} = -\begin{pmatrix} \ln\alpha_0 & \gamma & \cdots & \gamma \\ \gamma & \ln\alpha_1 & \cdots & \gamma \\ \vdots & \vdots & \ddots & \vdots \\ \gamma & \gamma & \cdots & \ln\alpha_{K-1} \end{pmatrix}. \tag{16}$$

The $N\times N$ matrix $\boldsymbol{H}(y_i)$ can be rewritten as

$$\begin{aligned} \boldsymbol{H}(y_i) &\equiv -\sum_{k=0}^{K-1}\sum_{k'=0}^{K-1} B_{kk'}^{(i)} \boldsymbol{X}_{kk'} \\ &= -\begin{pmatrix} \ln(\alpha_0 g_0(y_i)) & \gamma & \cdots & \gamma \\ \gamma & \ln(\alpha_1 g_1(y_i)) & \cdots & \gamma \\ \vdots & \vdots & & \vdots \\ \gamma & \gamma & \cdots & \ln(\alpha_{K-1} g_{K-1}(y_i)) \end{pmatrix}, \end{aligned} \tag{17}$$

where $\boldsymbol{X}_{k,k'}$ is a $K\times K$ matrix whose (l,l')-component $\langle l|\boldsymbol{X}_{kk'}|l'\rangle$ $(l,l' = 0,1,\cdots,K-1)$ are defines by

$$\langle l|\boldsymbol{X}_{kk'}|l'\rangle \equiv \delta_{k,l}\delta_{k',l'}, \tag{18}$$

and the coefficients $B_{k,k'}$ $(k, k' = 0, 1, \cdots, K-1)$ are defined by

$$B_{kk}^{(i)} \equiv \ln(\alpha_k g_k(\boldsymbol{y}_i)), \qquad B_{kk'}^{(i)} \equiv \gamma \ (k \neq k'). \tag{19}$$

The function $\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ defined by equations (14) and (17) does not always satisfy the normalization condition $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})dy_0dy_1\cdots dy_{N-1}=1$, so that it is not regarded as a probability density function of data set $\boldsymbol{y}$ In the present paper, we regard the function $\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ as a marginal likelihood approximately and formulate a quantum-mechanical extension of Gaussian mixture model in the present paper.

By using the linear response theory, we have

$$\frac{\partial}{\partial B_{kk'}^{(i)}}\ln(\mathrm{tr}\, e^{-H(y_i)}) = \frac{1}{\mathrm{tr}\, e^{-H(y_i)}}\mathrm{tr}\int_0^1 e^{-(1-\lambda)H(y_i)}X_{kk'}e^{-\lambda H(y_i)}d\lambda = \frac{\mathrm{tr}\, X_{kk'}e^{-H(y_i)}}{\mathrm{tr}\, e^{-H(y_i)}}. \tag{20}$$

The extremum conditions for the marginal likelihood $\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ with respect to $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ can be derived as

$$\mu_k = \frac{\sum_{i=0}^{N-1} y_i(\frac{\mathrm{tr}X_{kk}e^{-\boldsymbol{H}(y_i)}}{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}})}{\sum_{i=0}^{N-1}(\frac{\mathrm{tr}X_{kk}e^{-\boldsymbol{H}(y_i)}}{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}})}, \tag{21}$$

$${\sigma_k}^2 = \frac{\sum_{i=0}^{N-1}(y_i-\mu_k)^2(\frac{\mathrm{tr}X_{kk}e^{-\boldsymbol{H}(y_i)}}{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}})}{\sum_{i=0}^{N-1}(\frac{\mathrm{tr}\boldsymbol{X}_{kk}e^{-\boldsymbol{H}(y_i)}}{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}})}, \tag{22}$$

$$\alpha_k = \exp\Big(\mathrm{tr}\, \boldsymbol{X}_{k,k}\ln(\frac{1}{N}\sum_{i=0}^{N-1}\frac{e^{-\boldsymbol{H}(y_i)}}{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}})\Big). \tag{23}$$

For given data $\boldsymbol{y}$, we obtain estimates of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ by means of the iteration method numerically. We remark that equations (21)-(23) can be reduced to equations (6)-(8) by setting $\gamma = 0$.

When we introduce the energy matrix $\boldsymbol{H}(y_i)$, the posterior distribution $\mathcal{P}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ can be also replaced by

$$\boldsymbol{P}(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha}) = \frac{e^{-\boldsymbol{E}}}{\mathrm{tr}\, e^{-\boldsymbol{E}}}. \tag{24}$$

$$\begin{aligned}\boldsymbol{E} \equiv\ & \boldsymbol{H}(y_0)\otimes\boldsymbol{I}\otimes\boldsymbol{I}\otimes\cdots\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\boldsymbol{H}(y_1)\otimes\boldsymbol{I}\otimes\cdots\otimes\boldsymbol{I} \\ & +\boldsymbol{I}\otimes\boldsymbol{I}\otimes\boldsymbol{H}(y_2)\otimes\cdots\otimes\boldsymbol{I} + \cdots + \boldsymbol{I}\otimes\boldsymbol{I}\otimes\boldsymbol{I}\otimes\cdots\otimes\boldsymbol{H}(y_{N-1})\end{aligned} \tag{25}$$

The estimates of label for y_i can be given as an K-dimensional eigenvector $\widehat{\boldsymbol{x}}_i = (\widehat{x}_i^{(0)}, \widehat{x}_i^{(1)}, \cdots, \widehat{x}_i^{(K-1)})^{\mathrm{t}}$ which corresponds to the minimal eigenvalue of $K\times K$ local energy matrix $\boldsymbol{H}(y_i)$. Here all eigenvectors should be normalized in their absolute values. The eigenvector which corresponds to the minimal eigenvalue of $KN\times KN$ global energy matrix $\boldsymbol{E}$ is given by a KN-dimensional vector

$$\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{x}}_0\otimes\widehat{\boldsymbol{x}}_1\otimes\widehat{\boldsymbol{x}}_2\otimes\cdots\otimes\widehat{\boldsymbol{x}}_{N-1}. \tag{26}$$

4. Numerical Experiments

In this section we give some numerical experiments for estimating the hyperparameters $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ when the data $\boldsymbol{y}$ is set to a standard image. The obtained set of label can be regarded as a segmented image for a given standard image.

By using equations (21)-(23), we do some numerical experiments for a monochrome image given in figure 1(a). In table 1, we give estimates of hyperparameters, $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\sigma}}$ and the values of logarithm of marginal likelihood, $\mathcal{L}(\boldsymbol{y},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}}) \equiv \frac{1}{N}\ln\mathcal{P}(\boldsymbol{y}|\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}})$ for $K=3$. For the estimates, $\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}$, $\boldsymbol{\mu}=\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{\sigma}=\widehat{\boldsymbol{\sigma}}$, eigenvectors $\widehat{\boldsymbol{x}}_i=(\widehat{x}_i^{(0)},\widehat{x}_i^{(1)},\widehat{x}_i^{(2)})^{\rm t}$ which corresponds to the minimal eigenvalue of 3×3 matrix $\boldsymbol{H}(y_i)$ are drawn in figure 1(b)-(c). In figure 1(b)-(c), $\widehat{\boldsymbol{x}}_i=(1,0,0)^{\rm t}$, $\widehat{\boldsymbol{x}}_i=(0,1,0)^{\rm t}$ and $\widehat{\boldsymbol{x}}_i=(0,0,1)^{\rm t}$ correspond to red, green and blue, respectively.

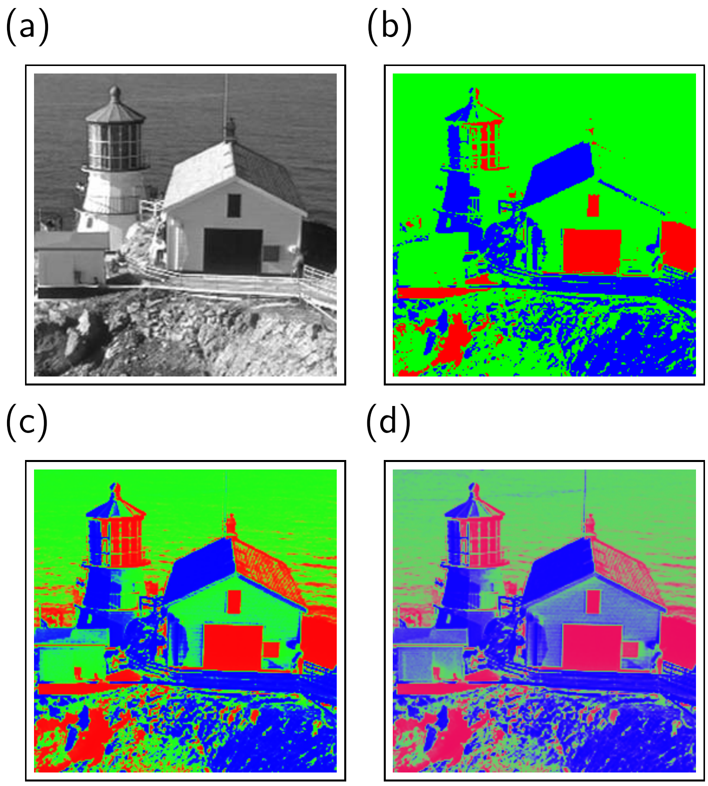

Figure 1. Image segmentation based on the quantum Gaussian mixture model for $K=3$. (a) Original image (256 grades, $N=256\times256$) (b) Segmented image $\widehat{\boldsymbol{x}}$ for $\gamma=0$. (c) Segmented image $\widehat{\boldsymbol{x}}$ for $\gamma=0.2$. (d) Segmented image $\widehat{\boldsymbol{x}}$ for $\gamma=0.4$.

The probability density function $\mathcal{P}(y_i|\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma}) \equiv {\rm tr}\, e^{-\boldsymbol{H}(y_i)}/{\rm tr}\, e^{-\boldsymbol{F}}$ and histogram for the given image $\boldsymbol{y}$ are shown in figure 2. In the curves of figure 2, the values of γ are set to 0, 0.2, and 0.4, respectively. In $\gamma=0.2$, we can find a very soft peak at $y_i=72$ around. The soft peak include some regions which are segmented as red areas for $\gamma=0.2$ and are segmented as green areas.

5. Concluding Remarks

In this paper, we have given an extension of the probability density function for conventional Gaussian mixture model to a density matrix. The formulation is based on Bayesian statistics and the maximum likelihood method. In the estimation of hyperparameters, we have to derive

Table 1. Estimates of hyperparameters, $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\sigma}}$ and values of $\mathcal{L}(\boldsymbol{y},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}}) \equiv \frac{1}{N}\ln\mathcal{P}(\boldsymbol{y}|\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}})$.

γ	$\mathcal{L}(\boldsymbol{y},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}})$		$\widehat{\alpha}_k$	$\widehat{\mu}_k$	$\widehat{\sigma}_k$
		$k=0$	0.0879	36.03	9.08
0	-5.3036	$k=1$	0.6526	107.17	24.76
		$k=2$	0.2595	195.33	36.55
		$k=0$	0.2679	72.62	32.94
0.2	-5.3492	$k=1$	0.3875	110.92	18.17
		$k=2$	0.3267	181.42	43.58
		$k=0$	0.2251	87.37	39.76
0.4	-5.3814	$k=1$	0.2311	107.53	17.29
		$k=2$	0.4437	152.11	58.59

the extremum conditions and we employee the linear response formulas for quantum-statistical models. We have constructed the estimation algorithm as an iterative procedure and have given some numerical experiments.

The function $\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ defined by equations (14) and (17) does not always satisfy the normalization condition so that it is not regarded as a probability density function of data set $\boldsymbol{y}$ In order to guarantee the normalization condition, we have to define

$$\mathcal{P}(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha}) = \prod_{i=0}^{N-1} \frac{\mathrm{tr}\, e^{-\boldsymbol{H}(y_i)}}{\int_{-\infty}^{+\infty} \mathrm{tr}\, e^{-\boldsymbol{H}(z_i)} dz_i}, \tag{27}$$

instead of equation (14). We should formulate our quantum-mechanical extension for equation (27). However, it may be difficult to calculate the integrals $\int_{-\infty}^{+\infty}\mathrm{tr}\, e^{-\boldsymbol{H}(z_i)}dz_i$ analytically. It remains as a future problem.

Density matrices include some quantum effects and are based on states constructed by superposing some kinds of classical states. For example, when we denote three possible classical states in terms of three-dimensional vectors $(1,0,0)^{\mathrm{t}}$, $(0,1,0)^{\mathrm{t}}$ and $(0,0,1)^{\mathrm{t}}$, all possible quantum states are expressed in terms of linear combinations of the three vectors in every extension to density matrix. In some capacities, quantum statistical models are expected to be beyond the conventional statistical models. For example, quantum statistical models may give us new optimal solutions in statistical inferences. In fact, our numerical experiments yields some nice segmentation results by introducing quantum states in the Gaussian mixture model. Quantum Gaussian mixture model has succeeded in splitting some regions from the background, though the conventional Gaussian mixture model cannot split the corresponding regions from the background.

In the present problems in this paper, we assume not to include interactions between any pairs of components y_i and y_j in given data $\boldsymbol{y} = (y_0, y_1, \cdots, y_{N-1})^{\mathrm{t}}$. Moreover we do not consider interactions between any pairs of elements in the set of labels. Thus the density matrix in equation (14) has been factorized with respect to each i. Though the energy matrix $\boldsymbol{H}$ in equation (25) seems to be an $NK\times NK$ matrix, the rank is just K. Such problems can be regarded as one-body problems in the statistical mechanics. If we consider interactions between some pairs of components, in the given data or in the set of label to estimate, it is hard to express the density matrix in terms of a factorizable form as shown in equation (14). When

Journal of Physics: Conference Series **95** (2008) 012023 doi:10.1088/1742-6596/95/1/012023

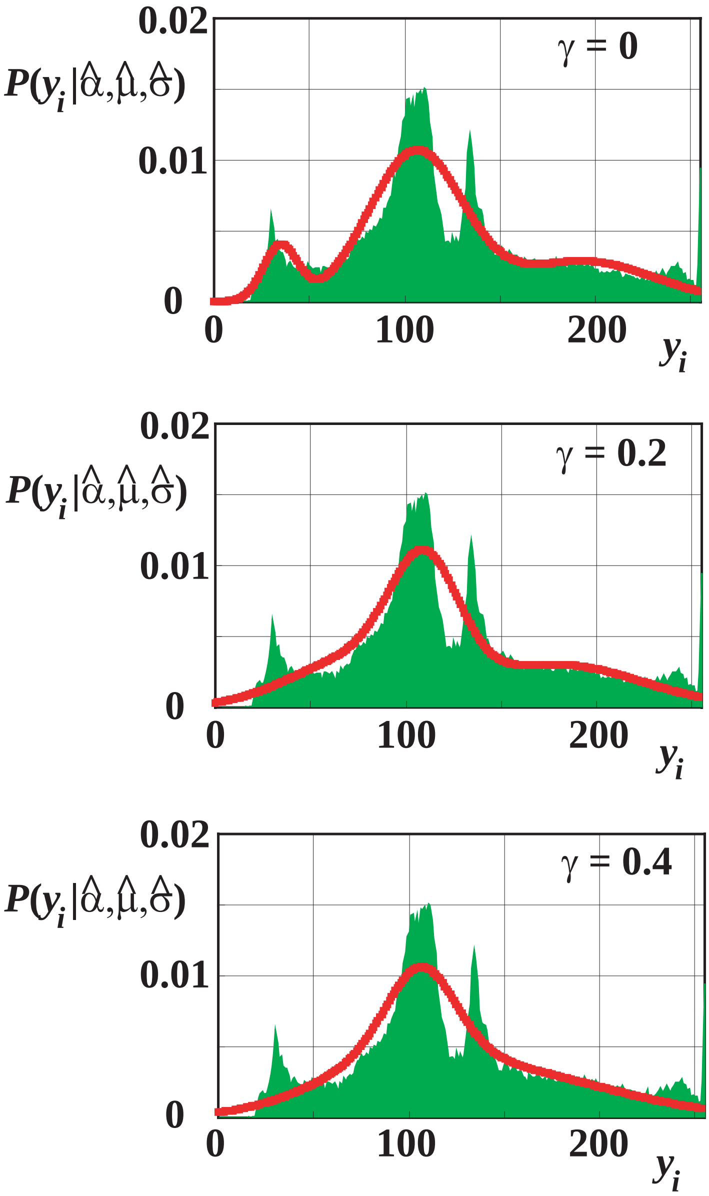

Figure 2. Curves of the probability density function $\mathcal{P}(y_i|\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma}) \equiv \mathrm{tr}\,e^{-\boldsymbol{H}(y_i)}/\mathrm{tr}\,e^{-\boldsymbol{F}}$ and histogram for the given image $\boldsymbol{y}$ in figure 1(a). The curves of red line show the values of $\mathcal{P}(y_i|\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\sigma}})$ for various values of y_i in the interval $[0,255]$. The hyperparameter γ is set to 0, 0.2 and 0.4. The histogram of the given image $\boldsymbol{y}$ in figure 1 is expressed in terms of green area in each graph.

the energy matrix $\boldsymbol{H}$ is given in such a way, the rank of $\boldsymbol{H}$ is not K any more and we have to diagonalize the $NK\times NK$ matrix $\boldsymbol{H}$. Many authors have investigated belief propagation and the other advanced mean-field methods to statistical inferences[15, 16, 17, 18, 19]. It is interesting to apply quantum belief propagation and the other advanced quantum mean-field methods to such cases as an approximate algorithm. Suzuki *et al.* investigated some quantum annealing algorithms by means of a quantum statistical version of Bethe approximation[20]. One of quantum statistical-mechanical extensions of belief propagations corresponds to a quantum cluster variation method[21]. It is interesting to apply a quantum cluster variation method to statistical inferences with quantum states and some interactions in the computer sciences. This is one of future problems.

Acknowledgements
This work was partly supported by the Grants-In-Aid (No.13680384, No.17500134 and No.18079002) for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Bishop C M 2006 *Pattern Recognition and Machine Learning* (New York: Springer)
[2] Ueda N and Nakano R 1998 *Neural Networks* **11** 271
[3] Kadowaki T and Nishimori H 1998 *Phys. Rev.* E **58** 5355
[4] Brooke J, Bitko D, Rosenbaum T F and Aeppli G 1999 *Science* **284** 779
[5] Santoro G E, Martoňák R, Tosatti E and Car R 2002 *Science* **295** 2427
[6] Santoro G E and Tossati E 2006 *J. Phys.* A: *Math. Gen.* **39** R393
[7] Suzuki S and Okada M 2005 *J. Phys. Soc. Jpn.* **74** 1649
[8] Tanaka K and Horiguchi T 1997 *IEICE Transactions* (A), **J80-A**, 2117 (in Japanese); translated in *Electronics and Communications in Japan*, *Part* 3: *Fundamental Electronic Science*, **83**, 84
[9] Morita S and Nishimori H 2006 *J. Phys.* A: *Math. Gen.* **39** 13903
[10] Dennis E, Kitaev A, Landahl A and Preskill J 2002 *J. Phys.* A: *Math. Gen.* **43** 4452
[11] Nishimori H and Sollich P 2004 *J. Phys. Soc. Jpn* **73** 2701
[12] Takeda K and Nishimori H 2004 *Nucl. Phys.* B **686** 377
[13] Tanaka K 2002 *J. Phys.* A: *Math. Gen.* **35** R81
[14] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing, —An Introduction—* (New York: Oxford University Press)
[15] Opper M and Saad D (eds) 2001 *Advanced Mean Field Methods — Theory and Practice —* (Cambridge: MIT Press)
[16] Tanaka K 2003 *IEICE Transactions on Information and Systems* **E86-D** 1228
[17] Ikeda S, Tanaka T and Amari S 2004 *Neural Computation* **16** 1779
[18] Yedidia J S, Freeman W T and Weiss Y 2005 *IEEE Transactions on Information Theory* **51** 2282
[19] Pelizzola A 2005 *J. Phys.* A: *Math. Gen.* **38** R309
[20] Suzuki S, Nishimori H and Suzuki M 2007 *Phys. Rev.* E **75** 051112
[21] Morita T 1957, *J. Phys. Soc. Jpn.* **12** 1060