IMAGE SEGMENTATION WITH MRF COUPLED INFINITE MIXTURE MODEL

Y. Cem Sübakan

Electrical & Electronics Engineering
Boğaziçi University, 34342 Bebek, Istanbul
{cem.subakan}@boun.edu.tr

ABSTRACT

Image segmentation is an important problem which addresses the needs of lots of biomedical applications. In this work, we adress the problem with MRF-coupled mixture models. In the standard finite mixture models, the number of segments that we are supposed to find in an image is fixed. With the Bayesian non-parametrics formulation we automatically find the number of segments by considering an infinite mixture model which consists of infinite number of clusters. We apply the MRF-coupled infinite mixture model on some biomedical images and show that with IMM, we are able to automatically determine the number of kinds of objects differing from the background.

Index Terms— Image Segmentation, Infinite Mixture Models, Markov Random Fields, Bayesian Non-Parametrics

1. INTRODUCTION

In image segmentation, the goal is to segment the objects in a scene as accurately as possible. One of the most prominent use of image segmentation algorithms are in biomedical applications. Generally the goal is to segment pathological parts from the rest so that the medical stuff can diagnose the status of the patient more easily and more accurately. In a particular application called lesion segmentation, the goal is to distinguish the slightly differing (intensity-wise) closed smooth regions from the background.

Most of the initial work are based on heuristics like thresholding and filtering. One example is [1]. These approaches clearly lack robustness due their lack of generality. There are also in-between work which can be dubbed quasi-heuristics which are based on algorithms such as region growing. We also have the probabilistic version. [94-98] Kupunski. Most of the initial probabilistic work is based on Markov random fields. However, they are generally maximum likelihood or at best MAP based, which are prone to under or overfitting. An example is [2]. There are some recent works which employ non-probabilistic machine learning techniques such as Boosting [3].

In the probabilistic image segmentation camp, to give several examples, there are works based on k-means clustering, Gaussian mixture models, histogram clustering [4, 5]. The main drawback of these works is that the models do not have spatial smoothness priors. Moreover, we have to specify the number of segments that is expected to be found a-priori. The problem of finding the number of clusters can be adressed by standard model selection. However this is an expensive and demanding process. An elegant way to solve this problem is to apply Non-parametric Bayesian methods. For example the infinite mixture model proposed in 2000, [6], is a clustering model where we automatically choose the number of clusters. The idea that follows is to couple infinite mixture model with the Markov random fields to obtain an image segmentation model which automatically finds the number of segments found within the image. This has been done in [7], which also handles infinite mixture models within the more general framework of Dirichlet processes. The goal of this work is to provide a simpler model for image segmentation using Bayesian Non-Parametrics paradigm.

The rest of the paper is organized as follows: In the following section we'll introduce the finite mixture models along with Markov Random Fields. Then we will talk on how to couple finite mixture models with MRFs. Finally we introduce Infinite mixture models. In the last section we provide results obtained using both models.

2. METHODOLOGY

2.1. Notation

Some of the most frequent notations used in the paper are as follows:

- x_{ij}; element in the i'th row and j'th column
- $x_{1:I}$ elements of x from 1 to I.
- $x_{1:I,1:J}$; elements in the row range 1 to I and column range 1 to J.
- $x^{-ij}_{1:I,1:J}$; elements in the row range 1 to I and column range 1 to J, except the element in the i'th row and j''th column.

- n_k^{-ij}; number of elements in cluster k except the element in the i''th row and j'th column.

In the following sections we'll introduce the models that make up to the ultimate segmentation model. We will describe the inference algorithms that are related with the ultimate segmentation model.

2.2. Finite Mixture Models (FMM)

For two dimensional data $x_{ij}, (i \in \{1 \dots I\}, \quad j \in \{1 \dots J\})$, a finite mixture model generates each data item x_{ij} from an observation density $F(x_{ij}; \theta_k)$ by randomly assigning them to the cluster k, where $k \in \{1 \dots K\}$ which is associated with parameters θ_k. The cluster assignments are done using the indicator variables z_{ij} which comes from a generic discrete distribution, Discrete$(z_{ij}; \pi)$. The parameters θ_k are drawn from the H distribution which is conjugate to $F(x_{ij}|\theta_k)$. (A non-conjugate distribution could also be chosen, however it would prevent the analytical derivations) The mixture proportions π_k comes from a Dirichlet distribution which is conjugate to the discrete distribution from which the indicator variables are drawn from. All in all, a standard, parametric finite mixture model can be defined as follows:

- Mixture proportions $\pi_{1:K}$

$$\pi|\alpha \quad \sim \quad \text{Dirichlet}(\alpha/K, \dots, \alpha/K) \tag{1}$$

- Parameters $\theta_{1:K}$

$$\theta_k|H \quad \sim \quad H \tag{2}$$

- Cluster indicator variables $z_{1:I,1:J}$

$$z_{ij}|\pi \quad \sim \quad \text{Discrete}(\pi) \tag{3}$$

- Data items $x_{1:I,1:J}$

$$x_{ij}|z_{ij}, \theta_k \quad \sim \quad F(\theta_k) \tag{4}$$

The corresponding directed graph is given in figure 1.

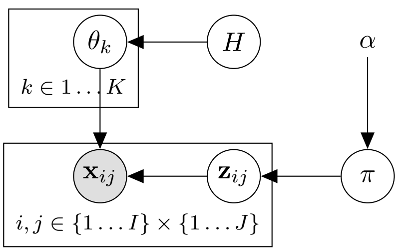

Fig. 1. The directed graphical model for a Finite Mixture Model

2.3. Markov Random Fields (MRF)

Markov Random Fields are flexible tools to impose a-priori constraints on structures like images. It basically models local interactions to impose a global constraint on the data. A markov random field is described using a Boltzmann distribution:

$$p(\mathbf{z}) = \frac{1}{Z} \exp\left(-E(\mathbf{z})\right) \tag{5}$$

where $E(\mathbf{z})$ is the energy function and defined as follows:

$$E(\mathbf{z}) = \sum_i \sum_j \underbrace{\sum_{m,n \in \mathcal{C}_{ij}} T(z_{ij}, z_{mn})}_{E(\mathbf{z}_{ij})} \tag{6}$$

This says us that the total energy, is dependent on the sum of local energies $E(\mathbf{z}_{ij})$, which are defined on the neighborhood system $\mathcal{C}_{ij}$. $T(z_{ij}, z_{mn})$ is a function that favors continuity in labels z_{ij}. For discrete inputs a and b, $T(a, b)$ is defined as;

$$T(a, b) = \begin{cases} -\beta & \text{if } a = b \\ \beta & \text{if } a \neq b \end{cases} \tag{7}$$

where β is a non-negative constant. So, if the label z_{ij} is identical to its neighbors, it increases the energy which increases the likelihood of its happening. Therefore, our purpose in using MRFs in the image segmentation model is to impose a continuity prior on labels.

2.4. FMM - MRF coupled model

The finite mixture model alone to segment an image would be a mere clustering of intensity value without taking into account any spatial information whatsoever. By assuming that, objects in an image form an closed, smooth surface, we enforce the labels to be spatially smoothly generated. A finite mixture model with the additional spatial constraints can be defined like a finite mixture model as defined in section 2.2, except that now, the indicator variables z_{ij} are not conditionally independent given π, but dependent on its neighbors in $\mathcal{C}_{ij}$.

$$p(z_{ij}|\pi, z_{mn}, (mn) \in \mathcal{C}_{ij}) \propto \text{Discrete}(\pi) \times \exp(-\sum_{mn \in C_{ij}} T(z_{ij}, z_{mn})) \tag{8}$$

A graphical model for this model can be constructed as shown in fig. 2.

2.4.1. *Inference with Gibbs sampling*

Gibbs sampling is a Markov Chain Monte-Carlo algorithm which draws samples from the posterior distribution of the

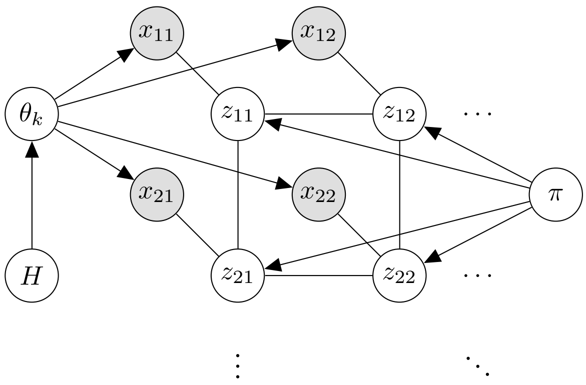

Fig. 2. The directed graphical model for the MRF-coupled finite mixture model

target variables. What we have to do is to write the full-joint distribution, and derive the full conditionals:

$$\begin{aligned}
&p(\pi, \theta_{1:K}, z_{1:I,1:J}, x_{1:I,1:J}) \qquad (9)\\
&= p(\pi) \prod_{k=1}^{K} p(\theta_k) \prod_{i=1,j=1}^{I,J} p(z_{ij}|z_{mn} \in \mathcal{C}_{ij}) \prod_{i=1,j=1}^{I,J} p(x_{ij}|z_{ij}, \theta_{1:K})\\
&= p(\pi) \prod_{k=1}^{K} H(\theta_k) \prod_{i=1,j=1}^{I,J} \prod_{k=1}^{K} \pi_k^{z_{ijk}} \exp(-\sum_{mn \in C_{ij}} T(z_{ij}, z_{mn}))\\
&\quad \times \prod_{i=1,j=1}^{I,J} \prod_{k=1}^{K} F(\theta_k)^{z_{ijk}}
\end{aligned}$$

Here z_{ij} is a vector which has only its k'th element 1 and else zero. Therefore it selects the k'th element. Then we derive the full conditionals for each variable by only considering the terms that have functional dependencies: The regular Gibbs sampler for the finite mixture model is as follows:

$$p(\theta_k|others) \propto H(\theta_k) \prod_{i,j:z_{ijk}=1} F(\theta_k) \qquad (10)$$

$$p(\pi|others) = \text{Dirichlet}(\alpha/K + n_1, \ldots, \alpha/K + n_K) \qquad (11)$$

$$p(z_{ijk} = 1|others) \propto \pi_k F(\theta_k) \exp(-\sum_{mn \in C_{ij}} T(z_{ij}, z_{mn})) \qquad (12)$$

2.4.2. *Inference with Collapsed Gibbs sampling*

In collapsed Gibbs sampling, we integrate out the variables that we are not interested in. Since this is an image segmentation algorithm, the important thing is to infer cluster indicator variables $z_{1:I,1:J}$. Therefore we integrate out the parameters $\theta_{1:K}$ and mixing proportions π. To derive the Gibbs sampler,

Algorithm 1 Regular Gibbs Sampling for FMM-MRF model

initialize $\theta_{1:K}$, π, $z_{1:I,1:J}$
for $\tau = 1 \to T$ **do**
  for $i = 1 \to I$ **do**
    for $j = 1 \to J$ **do**
      draw z_{ij}^{τ} from $p(z_{ij}|others)$
      draw θ_k^{τ} from $p(\theta_k|others)$.
      draw π^{τ} from $p(\pi|others)$.
    end for
  end for
end for

we have to derive the predictive full conditional

$$\begin{aligned}
&p(z_{ijk}|z_{1:I,1:J}^{-ij}, x_{1:I,1:J}) \propto \qquad (13)\\
&\qquad p(z_{ijk} = 1|\{z_{mn} : mn \neq ij, z_{mnk} = 1\})\\
&\qquad \times p(x_{ij}|\{x_{mn} : mn \neq ij, z_{mnk} = 1\})
\end{aligned}$$

We then write the necessary integration over π and $\theta_{1:K}$:

$$\begin{aligned}
&p(z_{ijk} = 1|z_{1:I,1:J}^{-ij}, x_{1:I,1:J}) \qquad (14)\\
&\propto \int \frac{\Gamma(\alpha)}{\prod_k \Gamma(\alpha/K)} \prod_{k=1}^{K} \pi_k^{\alpha/K-1} \prod_{k=1}^{K} \pi_k^{n_k} d\pi\\
&\quad \underbrace{\int \prod_{k=1}^{K} H(\theta_k) F(x_{ij}; \theta_k) \prod_{m,n \neq i,j:z_{ijk}=1} F(x_{mn}; \theta_k) d\theta_{1:K}}_{p(x_{ij}|\{x_{mn}:mn \neq ij, z_{mnk}=1\})}\\
&\quad \times \exp(-\sum_{ij} \sum_{mn \in C_{ij}} T(z_{ij}, z_{mn}))
\end{aligned} \qquad (15)$$

n_k is the number of data items assigned to cluster k. Note that to get the expression for predictive density $p(x_{ij}|\{x_{mn} : mn \neq ij, z_{mnk} = 1\})$ the observation model needs to be specified. After having the integrations worked out (See Appendix A for details), we get the following expression for the full conditional:

$$\begin{aligned}
&p(z_{ijk} = 1|z_{1:I,1:J}^{-ij}, x_{1:I,1:J}) \propto \qquad (16)\\
&\quad \frac{\alpha/K + n_k^{-i,j}}{\alpha + n - 1} p(x_{ij}|\{x_{mn} : mn \neq ij, z_{mnk} = 1\})\\
&\quad \times \exp(-\sum_{mn \in C_{ij}} T(z_{ij}, z_{mn}))
\end{aligned}$$

2.5. Infinite Mixture Model with MRF coupling

In infinite mixture model, we take the number of clusters to infinity. The easiest way to derive an inference algorithm for a MRF-coupled mixture model with an infinite number of clusters is to consider the collapsed Gibbs sampler we derived for

Algorithm 2 Collapsed Gibbs Sampling for FMM-MRF model

initialize $z_{1:I,1:J}$,
for $\tau = 1 \to T$ **do**
  for $i = 1 \to I$ **do**
    for $j = 1 \to J$ **do**
      draw z_{ij}^{τ} from $p(z_{ijk} = 1 | z_{1:I,1:J}^{-ij}, x_{1:I,1:J})$
    end for
  end for
end for

a MRF-coupled finite mixture model in section 2.4.2. What we do is to assume that the number of clusters K is a very large number and only K^* of the clusters are occupied. We lump all the empty clusters together. Then the full conditionals for the empty clusters turns out to be as follows:

$$p(z_{ijk} = 1, k = empty | z_{1:I,1:J}^{-ij}, x_{1:I,1:J}) \propto \frac{\alpha \frac{K-K^*}{K}}{\alpha + n - 1} p(x_{ij}) \times \exp(-4\beta) \quad (17)$$

What we essentially do to derive these three terms is to consider eq. (16), and apply it for the case where we assign a data item to an empty cluster. In this case since there is no member in that cluster, $n_k^{-i,j} = 0$, there is nothing to condition $p(x_{ij})$ to (See Appendix A to have the details on how obtain this marginal), and the all of the neighbors are different from the new guy, so the MRF coupling term returns a penalty of -4β. Then we take $K \to \infty$ to derive the collapsed Gibbs sampler for the infinite model. Full conditionals are as follows:

- Assignment to an already occupied cluster:

$$p(z_{ijk} = 1 | z_{1:I,1:J}^{-ij}, x_{1:I,1:J}) \propto \frac{n_k^{-i,j}}{\alpha + n - 1} p(x_{ij} | \{x_{mn} : mn \neq ij, z_{mnk} = 1\}) \times \exp(-\sum_{mn \in C_{ij}} T(z_{ij}, z_{mn})) \quad (18)$$

- Assignment to an empty cluster:

$$p(z_{ijk} = 1, k = empty | z_{1:I,1:J}^{-ij}, x_{1:I,1:J}) \propto \frac{\alpha}{\alpha + n - 1} p(x_{ij}) \times \exp(-4\beta) \quad (19)$$

3. RESULTS

In this section we provide some results obtained from the Infinite mixture model with the following observation model:

- Observation density:

$$F(x_{mn}; \theta_k) = \mathcal{N}(x_{mn}; \mu_k, \lambda^{-1}) \quad (20)$$

Algorithm 3 Collapsed Gibbs Sampling for IMM-MRF model

initialize $z_{1:I,1:J}$,
for $\tau = 1 \to T$ **do**
  for $i = 1 \to I$ **do**
    for $j = 1 \to J$ **do**
      draw z_{ij}^{τ} from $p(z_{ijk} = 1 | z_{1:I,1:J}^{-ij}, x_{1:I,1:J})$
      if $z_{ij(K+1)}^{\tau} = 1$ **then**
        $K = K + 1$
      end if
    end for
  end for
end for

- Prior:

$$H(\theta_k) = \mathcal{N}(\mu_k; \mu_0, (\kappa_0 \lambda)^{-1}) \mathcal{G}(\lambda; \alpha_0, \beta_0) \quad (21)$$

Some results on some 2D-liver scans are given in fig. 3.

In the medical image experiments, the parameter setting is: $\kappa_0 = 0.001, \alpha_0 = 10000, \beta_0 = 1000, \alpha = 0.001, \beta = 290$. We set μ_0 to the mean of the observed image. We started the experiments with $K = 2$. It finds $K = 3$ as expected. We see the advantage of using infinite mixture by seeing that the model not only segments dark stains, but also it is able to distinguish the other lighter regions from the background. In an another type of image, we obtain the result in fig. 3. We see that in the algorithm successfully segments the circular sign. It finds $K = 4$ segments as expected.

4. CONCLUSION AND FUTURE WORK

In this work, we have implemented collapsed Gibbs samplers for a MRF coupled finite mixture model and a MRF coupled infinite mixture model. We have seen that using infinite mixture enables us to automatically determine the number of segments which enables us to identify different types of objects present in the scene. In this work, we used two types of observation models: A Gaussian model with fixed variance and a fully Bayesian Gaussian observation model. Different observation models such as histogram clustering model or Poisson clustering model can also be implemented in future.

5. REFERENCES

[1] H. P. Chan, K. Doi, C. J. Vyborny, K. L. Lam, and R. A. Schmidt, "Computer-aided detection of microcalcifications in mammograms: methodology and preliminary clinical study," Investigative Radiology, vol. 23, no. 9, pp. 664-671, September 1988

[2] M. L. Comer, S. Liu, and E. J. Delp, "Statistical segmentation of mammograms," in Digital Mammography, K. Doi, Ed., International Congress Series. Amsterdam, the Netherlands: Elsevier, 1996, pp. 471?474.

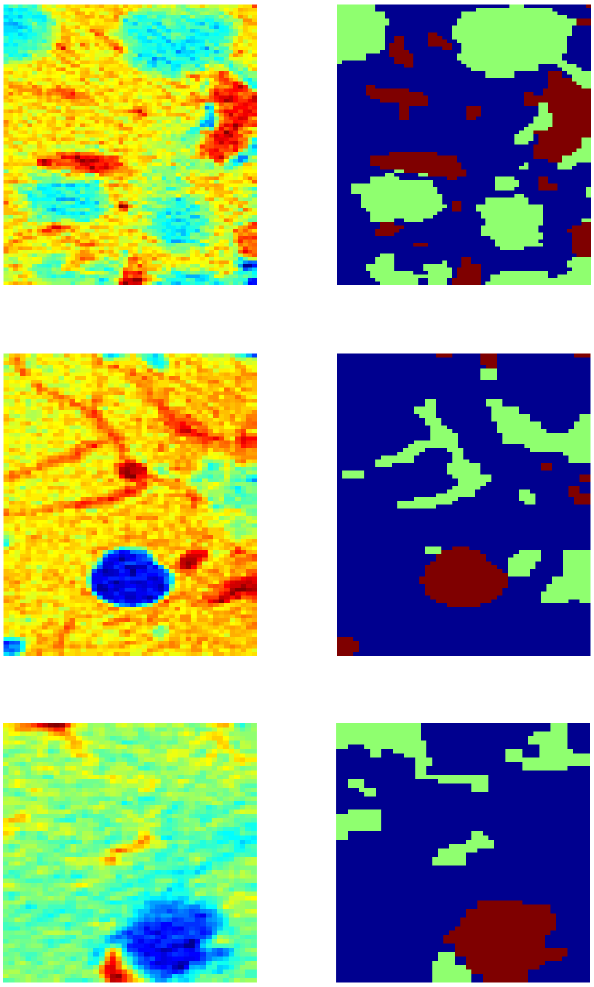

Fig. 3. Left column: The 2D liver scans, Right column: Segmented image

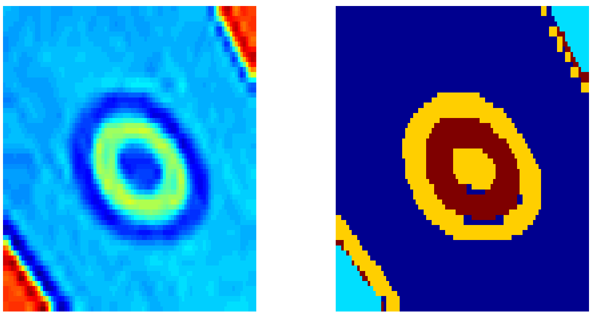

Fig. 4. Left column: Original Images, Right column: Segmented image

[3] Shimizu, T. Narihira, D. Furukawa, H. Kobatake, S. Nawano and K. Shinozaki, ”Ensemble segmentation using AdaBoost with application to liver lesion extraction from a CT volume”

[4] Samson, C., Blanc-Fraud, L., Aubert, G., and Zerubia, J. (2000). ”A variational model for image classi?cation and restoration” IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(5), 460?472.

[5] Hermes, L., Zller, T., and Buhmann, J. M. (2002). ”Parametric distributional clustering for image segmentation.” In Lecture notes in computer science: Vol. 2352. Computer vision ECCV ’02 (pp. 577?591). Berlin: Springer.

[6] C. E. Rasmussen. ”The infinite Gaussian mixture model”. In Advances in Neural Information Processing Systems, volume 12, 2000

[7] P.Orbanz, J.M.Buchmann, ”Nonparametric Bayesian Image Segmentation”, 2008, IJCV

[8] C.M. Bishop, Pattern Recognition and Machine Learning, 2006, Springer

[9] Murphy, Kevin P. (2007). ”Conjugate Bayesian analysis of the Gaussian distribution.”

6. APPENDIX A: OBSERVATION DENSITIES

6.1. Gaussian Observation Model, fixed variance

For occupied clusters, the integral for the predictive density is:

$$p(x_{ij}|\{x_{mn} : mn \neq ij, z_{mnk} = 1\}) = \int H(\theta_k)F(x_{ij};\theta_k) \prod_{m,n\neq i,j::z_{ijk}=1} F(x_{mn};\theta_k)d\theta_k \quad (22)$$

where, $H(\theta_k) = \mathcal{N}(\mu_k;\mu_0;\sigma_0^2)$ and $F(x_{mn};\theta_k) = \mathcal{N}(x_{mn};\mu_k,\sigma^2)$. Then, the procedure is to complete the squares to form a distribution in terms of θ_k. Having done that, this distribution integrates out to one, and we form another distribution with the remaining terms functionally related to x_{ij}. When we do this, the distribution obtained is as follows:

$$p(x_{ij}|\{x_{mn} : mn \neq ij, z_{mnk} = 1\}) = \mathcal{N}(x_{ij};m,v)$$

$$v = \frac{\sigma^2(\sigma^2 + (n_k^{-ij}+1)\sigma_0^2)}{\sigma^2 + n_k^{-ij}\sigma_0^2} \quad (23)$$

$$m = v\left[\left(\frac{\sum_{mn\neq ij:z_{mnk}=1} x_{ij}}{\sigma^4} + \frac{\mu_0}{\sigma_0^2\sigma^2}\right)\left(\frac{n_k^{-ij}}{\sigma^4} + \frac{1}{\sigma_0^2}\right)^{-1}\right]$$

$$= \frac{\sigma^2\mu_0 + \sigma_0^2 \sum_{mn\neq ij:z_{mnk}=1} x_{ij}}{\sigma^2 + n_k^{-ij}\sigma_0^2} \quad (24)$$

For empty clusters:

$$p(x_{ij}|\{x_{mn} : mn \neq ij, z_{mnk} = 1\}) \quad (25)$$

$$= \int H(\theta_k)F(x_{ij};\theta_k)d\theta_k$$

$$= \mathcal{N}(x_{ij};0;\sigma^2 + \sigma_0^2)$$

6.2. Gaussian Observation Model, Full Bayesian

For analytical convenience, we use precision instead of variance. The observation density is

$$F(x_{mn};\theta_k) = \mathcal{N}(x_{mn};\mu_k,\lambda^{-1})$$

. Prior is:

$$H(\theta_k) = \mathcal{N}(\mu_k;\mu_0,(\kappa_0\lambda)^{-1})\mathcal{G}(\lambda;\alpha_0,\beta_0)$$

, which is also known as Normal-Gamma density. For occupied clusters, the predictive density is:

$$p(x_{ij}|\{x_{mn}: mn \neq ij, z_{mnk}=1\}) = t(x_{ij};\mu_n,\Lambda,2\alpha_n) \quad (26)$$

$t(.)$ is known as student-t distribution. The parameters are;

$$\mu_n = \frac{\kappa_0\mu_0 + n_k^{-ij}}{\kappa_0 + n_k^{-ij}} \quad (27)$$

$$\Lambda = \frac{\alpha_n\kappa_n}{\beta_n(\kappa_n+1)} \quad (28)$$

$$\alpha_n = \alpha_0 + n_k^{-ij}/2 \quad (29)$$

$$\kappa_n = \kappa_0 + n/2 \quad (30)$$

$$\beta_n = \beta_0 + \frac{1}{2}\sum_{mn\neq ij: z_{mnk}=1}(x_{mn}-\bar{x})^2 + \frac{\kappa_0 n_k^{-ij}(\bar{x}-\mu_0)^2}{2(\kappa_0+n_k^{-ij})} \quad (31)$$

where, $\bar{x} = \sum_{mn\neq ij: z_{mnk}=1} x_{mn}/n_k^{-ij}$.

For empty clusters:

$$p(x_{ij}|\{x_{mn}: mn \neq ij, z_{mnk}=1\}) = t(x_{ij};\mu_0,\Lambda_0,2\alpha_0) \quad (32)$$

where $\Lambda_0 = \frac{\alpha_0\kappa_0}{\beta_0(\kappa_0+1)}$. Interested reader may refer to [9] for the detailed derivation.

7. APPENDIX B: DIRICHLET DISTRIBUTION

In this part we'll describe how to obtain marginal and predictive distributions for the indicator variables z_{ij}. The joint distribution of $z_{1:I,1:J}$ and π is:

$$p(\pi|\alpha)\times\prod_{i=1,j=1}^{I,J} p(z_{ij}|\pi) = \frac{\Gamma(\alpha)}{\prod_{k=1}^K\Gamma(\alpha/K)} \times\prod_{k=1}^K \pi_k^{\alpha/K-1}\times\prod_{k=1}^K\pi_k^{n_k} = \frac{\Gamma(\alpha)}{\prod_{k=1}^K\Gamma(\alpha/K)}\prod_{k=1}^K\pi_k^{\alpha/K+n_k-1} \quad (33)$$

Note that: $\prod_{k=1}^K \pi_k^{n_k} = \prod_{i=1,j=1}^{I,J}\prod_{k=1}^K \pi_k^{z_{ijk}}$. So, this is a Dirichlet distribution. To obtain the marginal $p(z_{1:I,1:J}|\alpha)$ is to multiply and divide this expression with proper normalization constant so that the distribution integrates out to one. We get:

$$p(z_{1:I,1:J}|\alpha) = \frac{\Gamma(\alpha)}{\prod_{k=1}^K\Gamma(\alpha/K)}\frac{\prod_{k=1}^K\Gamma(n_k+\alpha/K)}{\Gamma(n+\alpha)} \quad (34)$$

where $n=\sum_k n_k$. Then to obtain the predictive distribution $p(z_{ijk}=1|z_{1:I,1:J}^{-ij})$, we use the property $\Gamma(a+1)=a\Gamma(a)$:

$$\frac{\Gamma(\alpha)}{\prod_{k=1}^K\Gamma(\alpha/K)}\frac{\prod_{k=1}^K\Gamma(n_k+\alpha/K)}{\Gamma(n+\alpha)} = \underbrace{\frac{\Gamma(\alpha)}{\prod_{k=1}^K\Gamma(\alpha/K)}\frac{\prod_{k=1}^K\Gamma(n_k+\alpha/K-1)}{\Gamma(n+\alpha-1)}}_{p(z_{1:I,1:J}^{-ij})}\underbrace{\prod_{k=1}^K\frac{n_k^{-ij}+\alpha/K}{n+\alpha-1}}_{p(z_{ij}|z_{1:I,1:J}^{-ij})} \quad (35)$$

So, for $p(z_{ijk}=1|z_{1:I,1:J}^{-ij})$, we simply drop the multiplication over k.