VARIATIONAL POSTERIOR DISTRIBUTION APPROXIMATION IN BAYESIAN EMISSION TOMOGRAPHY RECONSTRUCTION USING A GAMMA MIXTURE PRIOR

Rafael Molina[1], Antonio López[2], José Manuel Martín[1] and Aggelos K. Katsaggelos[3]

[1]*Departamento de Ciencias de la Computación e I.A., Universidad de Granada, 18071 Granada, Spain*

[2]*Departamento de Lenguajes y Sistemas Informáticos, Universidad de Granada, 18071 Granada, Spain*

[3]*Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208-3118*

Keywords: Image reconstruction, parameter estimation, Bayesian framework, variational methods, Tomography images.

Abstract: Following the Bayesian framework we propose a method to reconstruct emission tomography images which uses gamma mixture prior and variational methods to approximate the posterior distribution of the unknown parameters and image instead of estimating them by using the Evidence Analysis or alternating between the estimation of parameters and image (Iterated Conditional Mode (ICM)) approach. By analyzing the posterior distribution approximation we can examine the quality of the proposed estimates. The method is tested on real Single Positron Emission Tomography (SPECT) images.

1 INTRODUCTION

SPECT (Single Photon Emission Computed Tomography) and PET (Positron Emission Tomography) are non invasive techniques which are used in Nuclear Medicine to take views of a isotope distribution in an patient. Since SPECT and PET obtain images via emission mode, both techniques are referred to as emission tomography.

In this paper, we address the problem of the reconstruction of emission tomography images. We propose the use of the hierarchical Bayesian framework to incorporate knowledge on the expected characteristics of the original image in the form of a mixture of gamma distributions, to model the observation process, and also to include information on the unknown parameters in the model in the form of hyperprior distributions. Then, by applying variational methods to approximate probability distributions we estimate the unknown parameters and the underlying original image.

The paper is organized as follows. In section 2 the Bayesian modeling and inference for our problem is presented. The used probability distributions for emission tomography images are formulated in section 3. The Bayesian analysis and posterior probability approximation to obtain the parameters and the original image is performed in section 4. The application of this method to a real SPECT study is described in section 5 and, finally, section 6 concludes the paper.

2 BAYESIAN FORMULATION

Let the object to be estimated be represented by a vector $\mathbf{x}$ of N lexicographically ordered voxels $\mathbf{x} = \{x_1,\ldots,x_N\}$. The observed, noisy data from which $\mathbf{x}$ is to be estimated is given by the vector $\mathbf{y}$, comprising lexicographically ordered elements $\mathbf{y} = \{y_1,\ldots,y_M\}$, where M is the number of detectors in the tomography system.

The Bayesian formulation of the Nuclear Medicine image reconstruction problem requires the definition of the joint distribution $\mathrm{p}(\boldsymbol{\Omega},\mathbf{x},\mathbf{y})$ of the observation $\mathbf{y}$, the unknown original image $\mathbf{x}$, and the hyperparameters $\boldsymbol{\Omega}$, describing their distributions.

To model the joint distribution we utilize the hierarchical Bayesian paradigm (see, for example (Molina et al., 1999; Galatsanos et al., 2002)). In the hierarchical approach we have two stages. In the first stage, knowledge about the structural form of the

Molina R., López A., Manuel Martín J. and K. Katsaggelos A. (2007).
VARIATIONAL POSTERIOR DISTRIBUTION APPROXIMATION IN BAYESIAN EMISSION TOMOGRAPHY RECONSTRUCTION USING A GAMMA MIXTURE PRIOR.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, pages 165-173
DOI: 10.5220/0002066001650173

observation noise and the structural behavior of the underlying image is used in forming $p(\mathbf{y} \mid \mathbf{x}, \Omega)$ and $p(\mathbf{x} \mid \Omega)$, respectively. In the second stage a hyperprior on the hyperparameters is defined, thus allowing the incorporation of information about these hyperparameters into the process.

Then the following joint distribution is defined for Ω, $\mathbf{x}$, and $\mathbf{y}$,

$$p(\Omega, \mathbf{x}, \mathbf{y}) = p(\Omega)p(\mathbf{x} \mid \Omega)p(\mathbf{y} \mid \mathbf{x}, \Omega), \quad (1)$$

and inference is based on $p(\Omega, \mathbf{x} \mid \mathbf{y})$ (see (Mohammad-Djafari, 1995), (Mohammad-Djafari, 1996)).

We can alternate the maximization of $p(\Omega, \mathbf{x} \mid \mathbf{y})$ with respect to Ω and $\mathbf{x}$ (the ICM approach), (Hsiao et al., 2002). However, this alternative maximization does not take into account the uncertainty in the original image when estimating the unknown parameters of the model and the consequential effect on the estimation of these parameters. An alternative methodology consists of estimating the hyperparameters in Ω by using

$$\hat{\Omega} = \arg\max_{\Omega} p(\Omega \mid \mathbf{y}) = \arg\max_{\Omega} \int_{\mathbf{x}} p(\Omega, \mathbf{x}, \mathbf{y}) d\mathbf{x}, \quad (2)$$

and then estimating the original image by solving

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x} \mid \hat{\Omega}, \mathbf{y}). \quad (3)$$

This inference procedure (called Evidence Analysis) aims at optimizing a given function and not at obtaining a posterior distribution that can be simulated to obtain additional information on the quality of the estimates.

The calculation of $p(\Omega, \mathbf{x} \mid \mathbf{y})$, however, may not be possible, in which case we have to decide how to approximate it. The Laplace approximation of distributions has been used, for instance, in blind deconvolution problems when the blur is partially known (Galatsanos et al., 2002; Galatsanos et al., 2000). An alternative method is provided by variational distribution approximation. This approximation can be thought of as being between the Laplace approximation and sampling methods (Andrieu et al., 2003). The basic underlying idea is to approximate $p(\Omega, \mathbf{x} \mid \mathbf{y})$ with a simpler distribution, usually one which assumes that $\mathbf{x}$ and the hyperparameters are independent given the data (see chapter II in (Beal, 2003) for an excellent introduction to variational methods and their relationships to other inference approaches).

The last few years have seen a growing interest in the application of variational methods (Likas and Galatsanos, 2004; Miskin, 2000; Molina et al., 2006) to inference problems. These methods attempt to approximate posterior distributions with the use of the Kullback-Leibler cross-entropy (Kullback, 1959). Application of variational methods to Bayesian inference problems include graphical models and neuronal networks (Jordan et al., 1998), independent component analysis (Miskin, 2000), mixture of factor analyzers, linear dynamic systems, hidden Markov models (Beal, 2003), support vector machines (Bishop and Tipping, 2000) and blind deconvolution problems (Miskin and MacKay, 2000; Likas and Galatsanos, 2004; Molina et al., 2006).

3 HYPERPRIORS, PRIORS AND OBSERVATION MODELS

For emission tomography the conditional distribution of the observed data $\mathbf{y}$ given $\mathbf{x}$ has the form

$$p(\mathbf{y} \mid \mathbf{x}) \propto \prod_{i=1}^{M} \exp\{-\sum_{j=1}^{N} A_{i,j}x_j\}(\sum_{j=1}^{N} A_{i,j}x_j)^{y_i}, \quad (4)$$

where $A_{i,j}$ is the contribution of the jth element of $\mathbf{x}$ to the ith element of $\mathbf{y}$. The system matrix A, with elements $A_{i,j}$, $i = 1, \ldots, M$, $j = 1, \ldots, N$ depends on the geometry of the gamma camera and effects, such as, the photon attenuation and the scatter contribution. This model together with the image model constitute the first stage of the hierarchical Bayesian modeling.

For the image to be estimated we use as prior model

$$p(\mathbf{x} \mid \pi, \beta, \boldsymbol{\alpha}) = \prod_{j=1}^{N} \left(\sum_{c=1}^{C} \pi_c p_G(x_j \mid \beta_c, \alpha_c) \right), \quad (5)$$

where, for a given number of classes C, π denotes the C-dimensional vector $\pi = (\pi_1, \ldots, \pi_C)$ consisting of C mixing proportions (weights) which are positive and satisfy the normalization constraint

$$\sum_{c=1}^{C} \pi_c = 1, \quad (6)$$

and $\boldsymbol{\alpha}$ and β denote, respectively, the C-dimensional vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_C)$, $\beta = (\beta_1, \ldots, \beta_C)$ satisfying $\alpha_c > 1$ and $\beta_c > 0$, $\forall c$. Each pair (α_c, β_c) defines for $x > 0$ the gamma probability distribution

$$p_G(x \mid \beta_c, \alpha_c) = \left(\frac{\alpha_c}{\beta_c} \right)^{\alpha_c} \frac{1}{\Gamma(\alpha_c)} x^{\alpha_c - 1} e^{-(\alpha_c/\beta_c)x}. \quad (7)$$

The mean, variance, and mode of this gamma distribution are given by

$$E[x] = \beta_c \quad Var[x] = \beta_c^2/\alpha_c \quad Mode[x] = \beta_c(1 - 1/\alpha_c). \quad (8)$$

The parameter β_c plays the role of the mean of cluster c while the pair β_c, α_c controls the variance of the prior distribution. There are then two nice interpretations of the parameter α_c it controls the smoothness of the reconstruction in class c and also measures the confidence on the prior mean. This second interpretation resembles the confidence values on the hyperprior parameters in image restoration problems (see, for instance, (Molina et al., 1999; Galatsanos et al., 2000; Galatsanos et al., 2002)). In this paper, following the approach in (Hsiao et al., 2002) we will not attempt to estimate α_c and leave it as an user-specified parameter.

The use of gamma priors in medical images was introduced in (Lange et al., 1987). To our knowledge the use of mixtures of gamma priors in medical imaging was first proposed in (Hsiao et al., 1998) for transmission tomography.

We now proceed to introduce the prior distribution (hyperprior) on the unknown parameters. We note that the set of unknown parameters is given by

$$\mathbf{\Omega} = \Big\{ \omega = (\pi, \beta) = (\pi_1, \ldots, \pi_C, \beta_1, \ldots, \beta_C) \mid \pi_c \geq 0 \; \forall c \text{ with } \sum_c \pi_c = 1, \text{ and } \beta_c > 0, \forall c \Big\}. \quad (9)$$

Following the Bayesian paradigm we have to define now the hyperprior distribution on $\mathbf{\Omega}$. We can use the following distribution on the unknown hyperparameters $\omega \in \mathbf{\Omega}$,

$$\mathrm{p}(\omega) = \mathrm{p}(\pi)\mathrm{p}(\beta), \quad (10)$$

where $\mathrm{p}(\pi)$ and $\mathrm{p}(\beta)$ are flat (assigning the same probability to all elements) distributions. We can, however, include additional information on the mixing weights by using as $\mathrm{p}(\pi)$ the C-variate Dirichlet distribution defined by

$$\begin{aligned} \mathrm{p}(\pi) &= \mathrm{p}(\pi_1, \ldots, \pi_C) \\ &= \frac{\Gamma(a_1 + \cdots + a_C)}{\Gamma(a_1) \cdots \Gamma(a_C)} \pi_1^{a_1 - 1} \cdots \pi_C^{a_C - 1}, \quad (11) \end{aligned}$$

over $\pi_c \geq 0 \; \forall c$, with $\sum_c \pi_c = 1$ and zero outside, where the a_c's are all real and positive. We will refer to a distribution having the density function given in equation (11) as the *C-variate Dirichlet distribution* $D(a_1, \ldots, a_C)$. A $D(a_1, \ldots, a_C)$ distribution has the following marginal means and variances,

$$\begin{aligned} E[\pi_c] &= \frac{a_c}{a_1 + \cdots + a_C} \\ Var[\pi_c] &= \frac{a_c(a_1 + \cdots + a_C - a_c)}{(a_1 + \cdots + a_C)^2 (a_1 + \cdots + a_C + 1)}, \\ & \quad c = 1, \ldots, C, \quad (12) \end{aligned}$$

(see (Wilks, 1962)). Note that if $a_c = \rho l_c$, where $l_c > 0$, $\forall c$ and $\rho > 0$, the mean of π_c does not depend on ρ, while ρ can be used to increase or decrease the variance of π_c.

We will assume that β_c, $c = 1, \ldots, C$ has as hyperprior distribution $p(\beta_c)$, the inverse gamma distribution defined by

$$\begin{aligned} \mathrm{p}_{IG}(\beta_c \mid m_c^0, n_c^0) &= \frac{((m_c^0 - 1) n_c^0)^{m_c^0}}{\Gamma(m_c^0)} \times \\ & \quad \beta_c^{-m_c^0 - 1} e^{-(m_c^0 - 1) n_c^0 / \beta_c}. \quad (13) \end{aligned}$$

where $m_c^0 > 1$ and $n_c^0 > 0$, $\forall c$, and the mean, variance, and mode of this inverse gamma distribution are given by

$$\begin{aligned} & E[\beta_c] = n_c^0, \quad Var[\beta_c] = (n_c^0)^2 / (m_c^0 - 2), \\ & Mode[\beta_c] = (m_c^0 - 1) n_c^0 / (m_c^0 + 1). \quad (14) \end{aligned}$$

We now have a probability distribution defined over $(\pi, \beta, \mathbf{x}, \mathbf{y})$ which has the form

$$\mathrm{p}(\pi, \beta, \mathbf{x}, \mathbf{y}) = \mathrm{p}(\pi)\mathrm{p}(\beta)\mathrm{p}(\mathbf{x} \mid \pi, \beta)\mathrm{p}(\mathbf{y} \mid \mathbf{x}) \quad (15)$$

4 BAYESIAN INFERENCE AND VARIATIONAL APPROXIMATION

In order to perform inference we need to either calculate or approximate the posterior distribution $\mathrm{p}(\pi, \beta, \mathbf{x} \mid \mathbf{y})$. Since this posterior distribution can not be found in closed form, we will apply variational methods to approximate this distribution by the distribution $\mathrm{q}(\pi, \beta, \mathbf{x})$.

The variational criterion used to find $\mathrm{q}(\pi, \beta, \mathbf{x})$ is the minimization of the Kullback-Leibler divergence, given by (Kullback and Leibler, 1951; Kullback, 1959)

$$\begin{aligned} & C_{KL}(\mathrm{q}(\pi, \beta, \mathbf{x}) \parallel \mathrm{p}(\pi, \beta, \mathbf{x} \mid \mathbf{y})) = \\ &= \int_{\pi, \beta, \mathbf{x}} \mathrm{q}(\pi, \beta, \mathbf{x}) \log \left(\frac{\mathrm{q}(\pi, \beta, \mathbf{x})}{\mathrm{p}(\pi, \beta, \mathbf{x} \mid \mathbf{y})} \right) d\pi d\beta d\mathbf{x} \\ &= \int_{\pi, \beta, \mathbf{x}} \mathrm{q}(\pi, \beta, \mathbf{x}) \log \left(\frac{\mathrm{q}(\pi, \beta, \mathbf{x})}{\mathrm{p}(\pi, \beta, \mathbf{x}, \mathbf{y})} \right) d\pi d\beta d\mathbf{x} \\ & \quad + \log \mathrm{p}(\mathbf{y}), \quad (16) \end{aligned}$$

which is always non negative and equal to zero only when $\mathrm{q}(\pi, \beta, \mathbf{x}) = \mathrm{p}(\pi, \beta, \mathbf{x} \mid \mathbf{y})$.

We choose to approximate the posterior distribution $\mathrm{p}(\pi, \beta, \mathbf{x} \mid \mathbf{y})$ by the distribution

$$\mathrm{q}(\pi, \beta, \mathbf{x}) = \mathrm{q}(\pi)\mathrm{q}(\beta)\mathrm{q}(\mathbf{x}), \quad (17)$$

where $\mathrm{q}(\pi)$, $\mathrm{q}(\beta)$ and $\mathrm{q}(\mathbf{x})$ denote distributions on π, β and $\mathbf{x}$ respectively. In the following we provide the derivation of two approximations of the posterior distribution as well as their algorithmic descriptions.

4.1 General Case

We now proceed to find the best of these distributions in the divergence sense.

Let

$$\Phi = \{\pi, \beta, \mathbf{x}\}. \tag{18}$$

For $\theta \in \Phi$ let us denote by Φ_θ the subset of Φ with θ removed; for instance, if $\theta = \mathbf{x}$, $\Phi_{\mathbf{x}} = (\pi, \beta)$. Then, Eq. (16) can be written as

$$\begin{aligned} & C_{KL}(\mathrm{q}(\pi,\beta,\mathbf{x}) \parallel \mathrm{p}(\pi,\beta,\mathbf{x} \mid \mathbf{y})) \\ = \;& C_{KL}(\mathrm{q}(\theta)\mathrm{q}(\Phi_\theta) \parallel \mathrm{p}(\pi,\beta,\mathbf{x} \mid \mathbf{y})) \\ = \;& \int_\theta \mathrm{q}(\theta) \left(\int_{\Phi_\theta} \mathrm{q}(\Phi_\theta) \log \left(\frac{\mathrm{q}(\theta)\mathrm{q}(\Phi_\theta)}{\mathrm{p}(\pi,\beta,\mathbf{x},\mathbf{y})} \right) d\Phi_\theta \right) d\theta \\ & + \text{const.} \end{aligned} \tag{19}$$

Now, given $\mathrm{q}(\Phi_\theta) = \prod_{\rho \neq \theta} \mathrm{q}(\rho)$, (if, for instance, $\theta = \mathbf{x}$ then $\mathrm{q}(\Phi_{\mathbf{x}}) = \mathrm{q}(\pi)\mathrm{q}(\beta)$), an estimate of $\mathrm{q}(\theta)$ is obtained as

$$\hat{\mathrm{q}}(\theta) = \arg\min_{\mathrm{q}(\theta)} C_{KL}\left(\mathrm{q}(\theta)\mathrm{q}(\Phi_\theta)\right) \parallel \mathrm{p}(\pi,\beta,\mathbf{x} \mid \mathbf{y})) . \tag{20}$$

Thus, we have the following iterative procedure to find $\mathrm{q}(\pi,\beta,\mathbf{x})$.

Algorithm 1 *General case. Iterative estimation of* $\mathrm{q}(\pi,\beta,\mathbf{x}) = \mathrm{q}(\pi)\mathrm{q}(\beta)\mathrm{q}(\mathbf{x})$.
Given $\mathrm{q}^1(\beta)$*, and* $\mathrm{q}^1(\mathbf{x})$*, the initial estimates of the distributions* $\mathrm{q}(\beta)$*, and* $\mathrm{q}(\mathbf{x})$*,*
For $k = 1, 2, \ldots$ *until a stopping criterion is met:*

1. *Find*

$$\mathrm{q}^k(\pi) = \arg\min_{\mathrm{q}(\pi)} C_{KL}(\mathrm{q}(\pi)\mathrm{q}^k(\beta)\mathrm{q}^k(\mathbf{x}) \parallel \mathrm{p}(\pi,\beta,\mathbf{x} \mid \mathbf{y})), \tag{21}$$

2. *Find*

$$\mathrm{q}^{k+1}(\beta) = \arg\min_{\mathrm{q}(\beta)} C_{KL}(\mathrm{q}^k(\pi)\mathrm{q}(\beta)\mathrm{q}^k(\mathbf{x}) \parallel \mathrm{p}(\pi,\beta,\mathbf{x} \mid \mathbf{y})), \tag{22}$$

3. *Find*

$$\mathrm{q}^{k+1}(\mathbf{x}) = \arg\min_{\mathrm{q}(\mathbf{x})} C_{KL}(\mathrm{q}^k(\pi)\mathrm{q}^{k+1}(\beta)\mathrm{q}(\mathbf{x}) \parallel \mathrm{p}(\pi,\beta,\mathbf{x} \mid \mathbf{y})) \tag{23}$$

Set

$$\mathrm{q}(\pi) = \lim_{k\to\infty} \mathrm{q}^k(\pi),\; \mathrm{q}(\beta) = \lim_{k\to\infty} \mathrm{q}^k(\beta),\; \mathrm{q}(\mathbf{x}) = \lim_{k\to\infty} \mathrm{q}^k(\mathbf{x}). \tag{24}$$

4.2 Degenerate Case

In the previous algorithm we have performed the search of the distributions $\mathrm{q}(\pi)$, $\mathrm{q}(\beta)$, and $\mathrm{q}(\mathbf{x})$ in an unrestricted manner. However, we can reduce the space of search to the set of degenerate distributions. This approach, to be developed now, will not provide information on the quality of the estimates but we use it to justify some of the estimation procedures proposed in the literature as the solution of the variational approach to posterior distributions when a particular distribution approximation is used.

Let

$$A = \left\{ \pi = (\pi_1, \ldots, \pi_C) \mid \; \pi_c \geq 0 \; \forall c \text{ with } \sum_c \pi_c = 1 \right\}. \tag{25}$$

and

$$B = \{\beta = (\beta_1, \ldots, \beta_C) \mid \; \beta_c > 0 \; \forall c \} . \tag{26}$$

Instead of using an unrestricted search for the distribution of $\mathrm{q}(\pi)$, $\mathrm{q}(\beta)$, and $\mathrm{q}(\mathbf{x})$ we will here restrict our search to the following sets of distributions

$$\begin{aligned} D(A) &= \{\mathrm{q}(\pi) \mid \; \mathrm{q}(\pi) \text{ is a degenerate} \\ & \quad\;\; \text{distribution on an element of A}\} && (27) \\ D(B) &= \{\mathrm{q}(\beta) \mid \; \mathrm{q}(\beta) \text{ is a degenerate} \\ & \quad\;\; \text{distribution on an element of B}\} && (28) \\ D((R_0^+)^N) &= \{\mathrm{q}(\mathbf{x}) \mid \; \mathrm{q}(\mathbf{x}) \text{ is a degenerate} \\ & \quad\;\; \text{distribution on an element of } (R_0^+)^N\}, && (29) \end{aligned}$$

where a degenerate distribution takes one value with probability one, that is,

$$\mathrm{q}(\theta) = \begin{cases} 1 & \text{if } \theta = \underline{\theta} \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

When $\mathrm{q}^k(\pi)$, $\mathrm{q}^k(\beta)$, and $\mathrm{q}^k(\mathbf{x})$ are degenerate distributions, we will denote by π^k, β^k, and $\mathbf{x}^k$ respectively the values these distributions take with probability one. We will also use the subscript D on the distributions $\mathrm{q}(\cdot)$ to denote the degenerate approximations $\mathrm{q}_D(\cdot)$. Then, the variational approach when using degenerate distributions becomes:

Algorithm 2 *Degenerate case. Iterative estimation of* $\mathrm{q}_D(\pi,\beta,\mathbf{x}) = \mathrm{q}_D(\pi)\mathrm{q}_D(\beta)\mathrm{q}_D(\mathbf{x})$
Given $\beta^1 \in B$*, and* $\mathbf{x}^1$*, the initial estimates of* β *and* $\mathbf{x}$*, respectively, for* $k = 1, 2, \ldots$ *until a stopping criterion is met:*

1. *Find*

$$\pi^k = \arg\min_{\pi \in A} \left\{ -\log \mathrm{p}(\pi, \beta^k, \mathbf{x}^k, \mathbf{y}) \right\}, \tag{31}$$

2. *Find*

$$\beta^{k+1} = \arg\min_{\beta \in B} \left\{ -\log \mathrm{p}(\pi^k, \beta, \mathbf{x}^k, \mathbf{y}) \right\}, \tag{32}$$

3. *Find*

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in(R_0^+)^N} \left\{ -\log \mathrm{p}(\pi^k, \beta^{k+1}, \mathbf{x}, \mathbf{y}) \right\} \tag{33}$$

Set

$$\mathrm{q}_D(\pi,\beta,\mathbf{x}) = \begin{cases} 1 & \textit{if } \pi = \lim_{k\to\infty}\pi^k, \beta = \lim_{k\to\infty}\beta^k, \\ & \mathbf{x} = \lim_{k\to\infty}\mathbf{x}^k \\ 0 & \textit{elsewhere} \end{cases} \tag{34}$$

Interestingly, this is the formulation used in (Hsiao et al., 2002) to estimate the hyperparameters and the image when flat distributions on π and β are used.

4.3 Implementation

In order to find the distributions solutions of algorithms 1 and 2, we define two sets of positive weights

$\Lambda = \{\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_N) \mid \boldsymbol{\lambda}_j = (\boldsymbol{\lambda}_{j,1}, ..., \boldsymbol{\lambda}_{j,C})$ satisfies $\sum_{c=1}^C \boldsymbol{\lambda}_{j,c} = 1$, $\boldsymbol{\lambda}_{j,c} \geq 0$, $c = 1, ..., C\}$
and
$\Upsilon = \{\boldsymbol{\mu} = (\boldsymbol{\mu}_1;\ j = 1...,N) \mid \boldsymbol{\mu}_j = (\boldsymbol{\mu}_{j,1}, ..., \boldsymbol{\mu}_{j,M})$ with $\sum_{j=1}^N \boldsymbol{\mu}_{j,i} = 1$, $\forall i\ \boldsymbol{\mu}_{j,i} \geq 0$, $\forall i, j\}$

Then for $\boldsymbol{\lambda} \in \Lambda$ and $\boldsymbol{\mu} \in \Upsilon$ we have

$$\begin{aligned} \log \mathrm{p}(\pi,\beta,\mathbf{x},\mathbf{y}) &= \log \mathrm{p}(\pi) + \log \mathrm{p}(\beta) \\ &+ \sum_j \log\left(\sum_{c=1}^C \pi_c \mathrm{p}_c(x_j \mid \beta_c, \boldsymbol{\alpha}_c)\right) - \sum_{i=1}^M \sum_{j=1}^N A_{i,j} x_j \\ &+ \sum_{i=1}^M y(i) \log\left(\sum_{j=1}^N A_{i,j} x_j\right) \\ &\geq \log \mathrm{p}(\pi) + \log \mathrm{p}(\beta) \\ &+ \sum_j \sum_{c=1}^C \boldsymbol{\lambda}_{j,c} \log\left(\frac{\pi_c}{\boldsymbol{\lambda}_{j,c}} \mathrm{p}_c(\mathbf{x}_j \mid \beta_c, \boldsymbol{\alpha}_c)\right) \\ &- \sum_{i=1}^M \sum_{j=1}^N A_{i,j} x_j + \sum_{i=1}^M y(i) \sum_{j=1}^N \boldsymbol{\mu}_{j,i} \log\left(\frac{A_{i,j}}{\boldsymbol{\mu}_{j,i}} x_j\right) \\ &= L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) \end{aligned} \tag{35}$$

In consequence, for $\boldsymbol{\lambda} \in \Lambda$ and $\boldsymbol{\mu} \in \Upsilon$ we have

$$-\log \mathrm{p}(\pi,\beta,\mathbf{x},\mathbf{y}) \leq -L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) \tag{36}$$

and

$$\int_\pi \int_\beta \int_{\mathbf{x}} \mathrm{q}(\pi,\beta,\mathbf{x}) \log\left(\frac{\mathrm{q}(\pi,\beta,\mathbf{x})}{\mathrm{p}(\pi,\beta,\mathbf{x},\mathbf{y})}\right) d\pi d\beta d\mathbf{x} \leq$$
$$\int_\pi \int_\beta \int_{\mathbf{x}} \mathrm{q}(\pi,\beta,\mathbf{x}) \log\left(\frac{\mathrm{q}(\pi,\beta,\mathbf{x})}{\exp[L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})]}\right) d\pi d\beta d\mathbf{x}$$

This leads to the following procedure to find the distributions $\mathrm{q}(\pi,\beta,\mathbf{x})$ or $\mathrm{q}_D(\pi,\beta,\mathbf{x})$. Note that we are summarizing the non-degenerate and degenerate cases in one algorithm.

Algorithm 3 *Iterative estimation of* $\mathrm{q}(\pi)$, $\mathrm{q}(\beta)$ *and* $\mathrm{q}(\mathbf{x})$ *or* $\mathrm{q}_D(\pi)$, $\mathrm{q}_D(\beta)$ *and* $\mathrm{q}_D(\mathbf{x})$.
Given $\mathrm{q}^1(\beta)$, $\mathrm{q}^1(\mathbf{x})$, *or* β^1 *and* $\mathbf{x}^1$ *and* $\boldsymbol{\lambda}^1 \in \Lambda$ *and* $\boldsymbol{\mu}^1 \in \Upsilon$
For $k = 1, 2, \ldots$ *until a stopping criterion is met:*

1. *Find the solution of*

$$\mathrm{q}^k(\pi) = \arg\min_{\mathrm{q}(\pi)} \Big(\int_\pi \int_\beta \int_{\mathbf{x}} \mathrm{q}(\pi)\mathrm{q}^k(\beta)\mathrm{q}^k(\mathbf{x}) \log\left(\frac{\mathrm{q}(\pi)\mathrm{q}^k(\beta)\mathrm{q}^k(\mathbf{x})}{\exp[L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda}^k,\boldsymbol{\mu}^k)]}\right) d\mathbf{x} d\beta d\pi \Big), \tag{37}$$

which is given by

$$\mathrm{q}^k(\pi) = \mathrm{p}_D(\pi \mid a_1 + \sum_{j=1}^N \lambda_{1,c}^k, \ldots, a_C + \sum_{j=1}^N \lambda_{1,C}^k), \tag{38}$$

or find the solution of

$$\pi^k = \arg\min_\pi \left\{ -L(\pi, \beta^k, \mathbf{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \right\} \tag{39}$$

which is given by

$$\pi_c^k = \frac{\sum_{j=1}^N \lambda_{j,c}^k + a_c - 1}{\sum_{c'=1}^C \sum_j \lambda_{j,c'}^k + \sum_{c'=1}^C (a_{c'} - 1)} \quad c = 1, \ldots, C \tag{40}$$

2. *Find the solution of*

$$\mathrm{q}^{k+1}(\beta) = \arg\min_{\mathrm{q}(\beta)} \Big(\int_\beta \int_\pi \int_{\mathbf{x}} \mathrm{q}^k(\pi)\mathrm{q}(\beta)\mathrm{q}^k(\mathbf{x}) \log\left(\frac{\mathrm{q}^k(\pi)\mathrm{q}(\beta)\mathrm{q}^k(\mathbf{x})}{\exp[L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda}^k,\boldsymbol{\mu}^k)]}\right) d\mathbf{x} d\pi d\beta \Big), \tag{41}$$

which is given by

$$\mathrm{q}^{k+1}(\beta) = \prod_{c=1}^C \mathrm{q}^{k+1}(\beta_c) \tag{42}$$

where

$$\mathrm{q}^{k+1}(\beta_c) = \mathrm{p}_{IG}\Big(\beta_c \mid m_c^0 + \alpha_c \textstyle\sum_j \lambda_{j,c}^k, (m_c^0 + \alpha_c \sum_j \lambda_{j,c}^k - 1) \frac{\alpha_c \sum_j \lambda_{j,c}^k E[x_j]_{q^k(\mathbf{x})} + (m_c^0 - 1) n_c^0}{m_c^0 + \alpha_c \sum_j \lambda_{j,c}^k - 1} \Big). \tag{43}$$

or find the solution of

$$\beta^{k+1} = \arg\min_\beta \left\{ -L(\pi^k, \beta, \mathbf{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \right\} \tag{44}$$

which is given by

$$\beta_c^{k+1} = \frac{\alpha_c \sum_j \lambda_{j,c}^k x_j^k + (m_c^0 - 1) n_c^0}{m_c^0 + \alpha_c \sum_j \lambda_{j,c}^k} \quad c = 1, \ldots, C \tag{45}$$

3. *Find the solution of*

$$q^{k+1}(\mathbf{x}) = \arg\min_{\mathrm{q}(\mathbf{x})} \Big(\int_{\mathbf{x}} \int_{\beta} \int_{\pi} \mathrm{q}^k(\pi)\mathrm{q}^{k+1}(\beta)\mathrm{q}(\mathbf{x}) \log\Big(\frac{\mathrm{q}^k(\pi)\mathrm{q}^{k+1}(\beta)\mathrm{q}(\mathbf{x})}{\exp[L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda}^k,\boldsymbol{\mu}^k)]} \Big) d\pi d\beta d\mathbf{x} \Big), \quad (46)$$

which is given by

$$\mathrm{q}^{k+1}(\mathbf{x}) = \prod_j \mathrm{q}^{k+1}(x_j) \quad (47)$$

where

$$\mathrm{q}^{k+1}(x_j) = \mathrm{p}_G(x_j \mid u_j^{k+1}, v_j^{k+1}) \quad (48)$$

and

$$u_j^{k+1} = \sum_c \lambda_{j,c}^k \alpha_c + \sum_i y(i)\mu_j^k \quad (49)$$

$$v_j^{k+1} = u_j^{k+1} / \left(\sum_c \lambda_{j,c}^k \alpha_c E\left[\frac{1}{\beta_c}\right]_{\mathrm{q}^{k+1}(\beta)} + \sum_i A_{i,j} \right) \quad (50)$$

or find the solution of

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \left\{ -L(\pi^k, \beta^{k+1}, \mathbf{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \right\} \quad (51)$$

which is given by

$$\mathbf{x}_j^{k+1} = (u_j^{k+1} - 1) / \left(\sum_c \lambda_{j,c}^k \frac{\alpha_c}{\beta_c^{k+1}} + \sum_i A_{i,j} \right) \quad (52)$$

4. *Find the solution of*

$$\boldsymbol{\lambda}^{k+1}, \boldsymbol{\mu}^{k+1} = \arg\min_{\boldsymbol{\lambda}\in\Lambda,\boldsymbol{\mu}\in\Upsilon} \Big(\int_{\pi} \int_{\beta} \int_{\mathbf{x}} \mathrm{q}^k(\pi)\mathrm{q}^{k+1}(\beta)\mathrm{q}^{k+1}(\mathbf{x}) \log\Big(\frac{\mathrm{q}^k(\pi)\mathrm{q}^{k+1}(\beta)\mathrm{q}^{k+1}(\mathbf{x})}{\exp[L(\pi,\beta,\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})]} \Big) d\pi d\beta d\mathbf{x} \Big), \quad (53)$$

which is given by

$$\lambda_{j,c}^{k+1} = \exp E\left[\log(\pi_c \mathrm{p}_G(x_j \mid \beta_c, \alpha_c))\right]_{q^{k+1}(\mathbf{x})\mathrm{q}^{k+1}(\beta)\mathrm{q}^k(\pi)} / \sum_{c'=1}^{C} \exp E\left[\log(\pi_{c'} \mathrm{p}_G(x_j \mid \beta_{c'}, \alpha_{c'}))\right]_{q^{k+1}(\mathbf{x})\mathrm{q}^{k+1}(\beta)\mathrm{q}^k(\pi)} \quad c = 1,\ldots,C \quad (54)$$

and

$$\boldsymbol{\mu}_{j,i}^{k+1} = \frac{\exp E\left[\log A_{i,j} x_j\right]_{q^{k+1}(\mathbf{x})}}{\sum_{j'=1}^{M} \exp E\left[\log A_{i,j'} x_{j'}\right]_{q^k(\mathbf{x})}} \quad j = 1,\ldots,N \quad (55)$$

or find the solution of

$$\boldsymbol{\lambda}^{k+1}, \boldsymbol{\mu}^{k+1} = \arg\min_{\boldsymbol{\lambda}\in\Lambda,\boldsymbol{\mu}\in\Upsilon} \left\{ -L(\pi^k, \beta^{k+1}, \mathbf{x}^{k+1}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \right\} \quad (56)$$

which is given by

$$\lambda_{j,c}^{k+1} = \frac{\pi_c^k \mathrm{p}_G(x_j^{k+1} \mid \beta_c^{k+1}, \alpha_c)}{\sum_{c'=1}^{C} \pi_{c'}^k \mathrm{p}_G(x_j^{k+1} \mid \beta_{c'}^{k+1}, \alpha_{c'})} \quad c = 1,\ldots,C \quad (57)$$

and

$$\boldsymbol{\mu}_{j,i}^{k+1} = \frac{A_{i,j} x_j^{k+1}}{\sum_{j'=1}^{M} A_{i,j'} x_{j'}^{k+1}} \quad j = 1,\ldots,N \quad (58)$$

For the non-degenerate case set

$$\mathrm{q}(\pi) = \lim_{k\to\infty} \mathrm{q}^k(\pi), \quad \mathrm{q}(\beta) = \lim_{k\to\infty} \mathrm{q}^k(\beta), \quad \mathrm{q}(\mathbf{x}) = \lim_{k\to\infty} \mathrm{q}^k(\mathbf{x}) \quad (59)$$

while for the degenerate case set

$$\mathrm{q}_D(\pi,\beta,\mathbf{x}) = \begin{cases} 1 & \textit{if } \pi = \lim_{k\to\infty} \pi^k, \beta = \lim_{k\to\infty} \beta^k \\ & \mathbf{x} = \lim_{k\to\infty} \mathbf{x}^k \\ 0 & \textit{elsewhere} \end{cases} \quad (60)$$

$E[x_j]_{q^k(x)}$, $E[\log(\pi_c p_G(x_j \mid \beta_c, \alpha_c))]_{q^{k+1}(x),q^{k+1}(\beta),q^k(\pi)}$, $E[1/\beta_c]_{q^{k+1}(\beta)}$ and $E[\log A_{i,j}x_j]_{q^{k+1}(x)}$ are calculated in the Appendix.

5 EXPERIMENTAL RESULTS

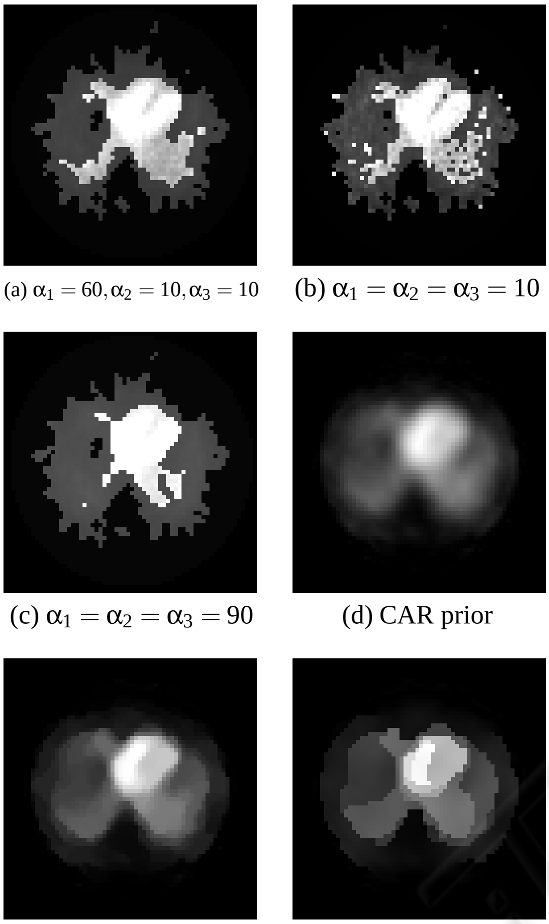
In order to evaluate the proposed method we have performed a set of tests over a real thoracic SPECT study. Emission images of a thorax present abrupt edges in the transition between tissues, therefore the gamma mixture prior is well adapted to the characteristics of these images.

The detector system used is a Siemens Orbiter 6601. The detector described a circular orbit clockwise, at 5.625 steps (there are 64 angles, 64 bins, and 64 slices). The data given by the detector system were corrected for the attenuation effect.

We have centered our attention in a cross sectional slice of the heart which presents a significant structure. This cut corresponds to the inferior part of the left ventricle of the heart and the superior area of the liver.

In the experiments we use three classes (background, liver and ventricle). Very similar results are obtained by the degenerate and non-degenerate posterior distribution approximations. However, the non-degenerate reconstructions are sharper and less noisy. Thus, we only present the non-degenerate results here.

The initial values $q^1(\beta)$, $q^1(\mathbf{x})$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ were estimated from a C-mean clustering of the Filtered Backprojection (FBP) reconstruction. The parameters α_1,



(a) $\alpha_1 = 60, \alpha_2 = 10, \alpha_3 = 10$ (b) $\alpha_1 = \alpha_2 = \alpha_3 = 10$

(c) $\alpha_1 = \alpha_2 = \alpha_3 = 90$ (d) CAR prior

(e) GGMRF prior (f) CGMRF prior

Figure 1: Results with SPECT image.

α_2, and α_3 were selected experimentally to obtain an acceptable visual tradeoff between detail and noise reduction. These parameters were equal to 60, 50, and 32, respectively. Figure 1(a) shows the reconstruction obtained by the proposed method. The ventricle is clear in the image and we can observe a small black area in the left region of the image (possible tumor). The patient presented a symptomatology compatible with a hepatic tumor. Small values of α_c produce very noisy reconstructions (see Fig. 1(b)), while large values of α_c cluster in excess the pixels of the reconstructed image (see Fig. 1(c), where part of the myocardial wall is not distinguishable).

The variances of the gamma distributions $q(\mathbf{x}_j) = q(x_j \mid u_j, v_j)$ provide information about the influence of the α_c's on the reconstruction. These variances are given by $var[x_j]_{q(x_j|u_j,v_j)} = v_j^2/u_j$. For our experiment, Fig. 2 shows the dependence of the mean of the above variances on α_1, α_2, and α_3 (we plot the curves that correspond to each α_c parameter with fixed values for the other two parameters). We can observe that the value of α_1 (background and tumor) is not especially critical for the reconstruction while we need a more careful selection of the parameters α_2 (liver) and α_3 (ventricle).

For visual comparison, since the original image is obviously not available, we show the obtained reconstruction using several image priors: conditional autorregresive (CAR), generalized Gauss Markov random fields (GGMRF) and compound Gauss Markov random field (CGMRF) (see Figs. 1(d), 1(e) and 1(f)), respectively). These reconstructions were obtained with the proposed methods in (López et al., 2004; López et al., 2002) The area of the tumor is clearer in the image provided by the reconstruction using a gamma mixture prior. For example, the lesion zone does not appear in the obtained reconstruction with the CGMRF prior (see Fig. 1(f)). We note that in the proposed method the isotope activity is homogeneous in the left ventricle, but it is not homogeneous in the reconstruction obtained by using the other priors.

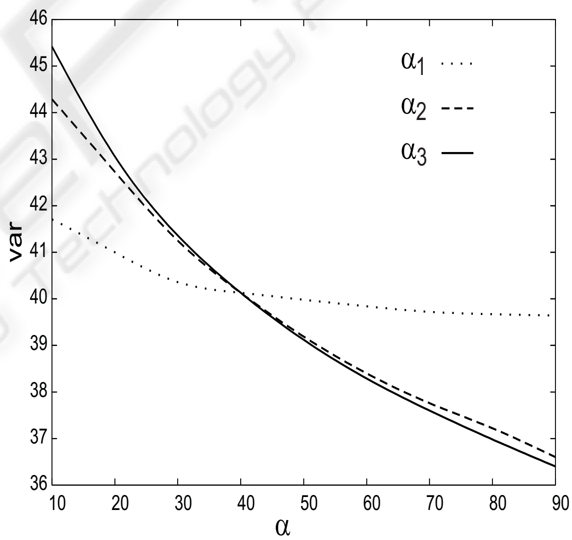


Figure 2: Mean of $var[x]$ with respect to $\boldsymbol{\alpha_c}$.

6 CONCLUSION

We present a reconstruction method for emission computed tomography which uses a gamma mixture as prior distribution to reconstruct Nuclear Medicine images that present abrupt changes of activity between contiguous tissues, since spatially independent priors, as the gamma mixture, are more adapted to this type of images. We use variational methods to obtain the original image and parameters estimation within an unified framework. Satisfactory experimental results are obtained with real clinical images.

ACKNOWLEDGEMENTS

This work has been partially supported by FIS projects PI04/0857, PI05/1953 and by the Greece-Spain Integrated Action HG2004-0014.

REFERENCES

Andrieu, C., de Freitras, N., Doucet, A., and Jordan, M. (2003). An introduction to MCMC for machine learning. *Machine Learning*, 50:5–43.

Beal, M. (2003). *Variational algorithms for approximate Bayesian inference*. PhD thesis, The Gatsby Computational Neuroscience Unit, University College London.

Bishop, C. and Tipping, M. (2000). Variational relevance vector machine. In *Proceedings of the 16th Conference on Uncertainty in Articial Intelligence*, pages 46–53. Morgan Kaufmann Publishers.

Galatsanos, N. P., Mesarovic, V. Z., Molina, R., and Katsaggelos, A. K. (2000). Hierarchical bayesian image restoration for partially-known blur. *IEEE Trans Image Process*, 9(10):1784–1797.

Galatsanos, N. P., Mesarovic, V. Z., Molina, R., Katsaggelos, A. K., and Mateos, J. (2002). Hyperparameter estimation in image restoration problems with partially-known blurs. *Optical Eng.*, 41(8):1845–1854.

Hsiao, I.-T., Rangarajan, A., and Gini, G. (1998). Joint-map reconstruction/segmentation for transmission tomography using mixture-models as priors. In *Proceedings of EEE Nuclear Science Symposium and Medical Imaging Conference*, volume II, pages 1689–1693.

Hsiao, I.-T., Rangarajan, A., and Gini, G. (2002). Joint-map Bayesian tomographic reconstruction with a gamma-mixture prior. *IEEE Trans Image Process*, 11:1466–1477.

Jordan, M. I., Ghahramani, Z., Jaakola, T. S., and Saul, L. K. (1998). An introduction to variational methods for graphical models. In *Learning in Graphical Models*, pages 105–162. MIT Press.

Kullback, S. (1959). *Information Theory and Statistics*. New York, Dover Publications.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Lange, K., Bahn, M., and Little, R. (1987). A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Trans Med Imag*, 6:106–114.

Likas, A. C. and Galatsanos, N. P. (2004). A variational approach for Bayesian blind image deconvolution. *ieee j sp*, 52(8):2222–2233.

López, A., Molina, R., Katsaggelos, A. K., Rodriguez, A., López, J. M., and Llamas, J. M. (2004). Parameter estimation in bayesian reconstruction of SPECT images: An aid in nuclear medicine diagnosis. *Int J Imaging Syst Technol*, 14:21–27.

López, A., Molina, R., Mateos, J., and Katsaggelos, A. K. (2002). SPECT image reconstruction using compound prior models. *Int J Pattern Recognit Artif Intell*, 16:317–330.

Miskin, J. (2000). *Ensemble Learning for Independent Component Analysis*. PhD thesis, Astrophysics Group, University of Cambridge.

Miskin, J. W. and MacKay, D. J. C. (2000). Ensemble learning for blind image separation and deconvolution. In Girolami, M., editor, *Advances in Independent Component Analysis*. Springer-Verlag.

Mohammad-Djafari, A. (1995). A full bayesian approach for inverse problems. In *in Maximum Entropy and Bayesian Methods, Kluwer Academic Publishers, K. Hanson and R.N. Silver eds. (MaxEnt95)*.

Mohammad-Djafari, A. (1996). Joint estimation of parameters and hyperparameters in a bayesian approach of solving inverse problems. In *Proceedings of the International Conference on Image Processing*, pages 473–477.

Molina, R., Katsaggelos, A. K., and Mateos, J. (1999). Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Trans Image Process*, 8(2):231–246.

Molina, R., Mateos, J., and Katsaggelos, A. K. (2006). Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans Image Process*, 15(12):3715–3727.

Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley and Sons.

APPENDIX

In algorithm 3 we need to calculate the quantities $E[x_j]_{q^k(x)}$, $E[\log(\pi_c p_G(x_j \mid \beta_c, \alpha_c))]_{q^{k+1}(x), q^{k+1}(\beta), q^k(\pi)}$, $E[1/\beta_c]_{q^{k+1}(\beta)}$ and $E[\log A_{i,j} x_j]_{q^{k+1}(x)}$.

To calculate $E[x_j]_{q^k(x)}$ we note that (see Eq. 8)

$$E[x_j]_{p_G(x_j|u_j,v_j)} = v_j \tag{61}$$

To calculate $E[1/\beta_c]_{q^{k+1}(\beta)}$ we observe with the use of Eq. 13 that

$$\begin{aligned} E\left[\frac{1}{\beta_c}\right]_{q(\beta_c|r_c,s_c)} &= \\ &= \int_{\beta_c} \frac{((r_c-1)s_c)^{r_c}}{\Gamma(r_c)} \beta_c^{-(r_c+1)-1} e^{-(r_c-1)s_c/\beta_c} d\beta_c \\ &= \frac{((r_c-1)s_c)^{r_c}}{\Gamma(r_c)} \frac{\Gamma(r_c+1)}{((r_c-1)s_c)^{r_c+1}} = \frac{r_c}{(r_c-1)s_c} \end{aligned} \tag{62}$$

We can easily calculate $E[\log A_{i,j}x_j]_{q^{k+1}(x)}$, since

$$E[\log x_j]_{p_G(x_j|u_j,v_j)} = -\log\frac{u_j}{v_j} + \frac{\partial \log\Gamma(u)}{\partial u}\ |_{u=u_j} \tag{63}$$

where

$$\Gamma(u) = \int_0^\infty \tau^{u-1}e^{-\tau}d\tau. \tag{64}$$

(see (Miskin, 2000)). $\partial\log\Gamma(u)/\partial u$ is the so called ψ or digamma function.

To calculate the expectation $E[\log(\pi_c p_G(x_j \mid \beta_c,\alpha_c))]_{q^{k+1}(x),q^{k+1}(\beta),q^k(\pi)}$ we note that

$$\begin{aligned} & E[\log(\pi_c p_G(x_j \mid \beta_c,\alpha_c))]_{q^{k+1}(x),q^{k+1}(\beta),q^k(\pi)} = \\ = \quad & E[\log\pi_c]_{q^k(\pi)} + \alpha_c\log\alpha_c - \alpha_c E[\log\beta_c]_{q^{k+1}(\beta)} \\ & -\log\Gamma(\alpha_c) + (\alpha_c-1)E[\log(x)]_{q^{k+1}(x)} \\ & -\alpha_c E[\frac{1}{\beta_c}]_{q_{k+1}(\beta)}E[x_j]_{q^{k+1}(x)} \end{aligned} \tag{65}$$

where $E[\log\pi_c]_{q^k(\pi)}$ can be calculated taking into account that

$$\begin{aligned} E[\log\pi_c]_{p_D(\pi_c|\omega,\ldots,\omega_c)} = & \frac{\partial\log\Gamma(\omega)}{\partial\omega}\ |_{\omega=\omega_c} \\ & -\frac{\partial\log\Gamma(\omega)}{\partial\omega}\ |_{\omega=\sum_{c'=1}^{C}\omega_{c'}} \end{aligned} \tag{66}$$

and $E[\log\beta_c]_{q^{k+1}(\beta_c)}$ can be calculated observing that β_c follows a distribution $p_G(\rho_c \mid m_c, 1/n_c)$ and since

$$E[\log\beta_c]_{p_{IG}(\beta_c|m_c,n_c)} = -E[\log\rho_c]_{p_G(\rho_c|m_c,1/n_c)} \tag{67}$$

which can be calculated using Eq. 63.