DETERMINISTIC APPROACH TO THE POSTERIOR DISTRIBUTION IN A BAYESIAN IMAGING MODEL

JIANG QIAN
VALEN JOHNSON
Institute of Statistics & Decision Sciences, Duke University
JAMES BOWSHER
RONALD JASZCZAK
Department of Radiology, Duke University

DP #92-A37
Deterministic approach to the posterior distribution in a Bayesian imaging model

Jiang Qian¹, Valen Johnson¹, James Bowsher², and Ronald Jaszczak²

¹Inst of Stat and Decision Sci., Duke University, Durham, NC 27708
²Department of Radiology, Duke University Medical Center, Durham, NC 27710

ABSTRACT

We describe a deterministic approach for approximating the posterior uncertainties of region volumes and means obtained from a statistical model for segmentation and analysis of emission computed tomography (ECT) images. Use of this approach in conjunction with Newton-Raphson iterative techniques and Laplace approximations to the posterior distribution on the image scene makes the resulting statistical model feasible for routine applications in PET and 3-D SPECT reconstructions.

1. INTRODUCTION

In Johnson 1991(c), a Bayesian model for image reconstruction and segmentation is presented. The Bayesian model is specified hierarchically within the Bayesian paradigm, and is parameterized in terms of quantities of direct physical interest. In particular, parameters representing intensity-differentiated regions and their means appear explicitly in the model. As shown in that paper, posterior uncertainties in region means and volumes can be assessed by Gibbs sampling technique. However in realistic reconstructions from projection data, the required forward and backward procedures are computationally demanding, and sampling from the posterior may not be practical. Thus appropriate approximations to the sampling distribution may be necessary for SPECT/PET.

The model is specified in stages. In the first stage, a probability distribution is placed on all partitions of the image space into connected sets. In the second stage, region means are associated with all possible partitioning sets. In the third stage, pixel intensities are assumed drawn from gamma distributions centered on the region means. Observations are generated according to Poisson distributions with means equal to linear combinations of the pixel intensities in the fourth stage.

2. MODEL DESCRIPTION

For completeness, details concerning model specification within each stage are reviewed below. Much of the description is taken from Johnson 1991(c).
2.1 The Prior Partition Model

A first assumption made in the model is that the true scene can be adequately represented by a discrete array of pixels, and that image intensity is constant within pixels. In most previous restoration/reconstruction models, pixels have been arranged on rectangular lattices to facilitate indexing, but we consider primarily hexagonal lattices. Hexagonal lattices provide more symmetric neighborhood structures for pixels and simplify the definition of connected regions. The pixel array is denoted by \( \Xi = \{ G_i \} \).

Another important assumption in our model is that the true scene consists of an unknown number of intensity-differentiated objects. Associated with each possible configuration of objects in the scene is a partition of \( \Xi \), where a partition is defined here as any collection of sets of connected pixels in which each pixel appears in one and only one set. A set of pixels will be considered connected if it is possible to move from any pixel in the set to any other pixel in the set without leaving the set. When \( \Xi \) is defined on a rectangular array, movement between pixels that either touch at a corner or share a common side is permitted, while for hexagonal arrays only movement between pixels that share a common side is permitted.

In order to define a probability distribution on the class of all partitions, assign to every pixel in the array \( \Xi \) an integer such that all pixels in each partitioning set are assigned the same integer, and each partitioning set is associated with a distinct integer. The particular integers chosen are otherwise arbitrary. These integers are called region identifiers and the array of region identifiers is denoted \( R = \{ r_i \} \).

The first stage model for the region identifiers \( R \) is specified as a Gibbs distribution. Gibbs distributions are in turn defined in terms of neighborhood systems, cliques, and potential functions. A neighborhood system on a graph \( \Xi \) is defined to be any collection of subsets \( G = \{ G_\xi, \xi \in \Xi \} \) such that \( \xi \not\in G_\xi \), and \( \xi_1 \in G_{\xi_2} \) if and only if \( \xi_2 \in G_{\xi_1} \). A clique is defined as any subset of \( \Xi \) in which every element is in the neighborhood of every other element (note that single pixels satisfy this requirement by default). Denote the set of cliques by \( C \).

A portion of a simple neighborhood system defined on a hexagonal lattice is depicted in Figure 1. The subgraph in Figure 1a shows the six first order neighbors of the central pixel. Up to arbitrary rotations, the three clique types that result from this neighborhood structure are depicted in Figure 1b.

Given the array \( \Xi \) and a neighborhood system \( G \), a Gibbs distribution on the region identifier array \( R \) must have a density function expressible in the form

\[
\pi(R) = \frac{1}{Z} \exp\{-U(R)\}, \quad U(R) = \sum_{C \in C} V_C(R).
\]

In this expression, \( Z \) is called the partition function and is a constant independent of \( R \). \( U(R) \) is called
the energy function, and the functions $V_C$ are called potentials. Importantly, the potential $V_C$ can depend only on components of $R$ that belong to the subscripted clique $C$. An important consequence of the form of the energy function is that the conditional distribution of any component in the system, say $r_i$, given the values of all other components of $R$, depends only on the potentials of cliques that contain site $r_i$. In other words, only cliques containing $r_i$ and their associated potentials need be considered when determining the conditional distribution of the site $r_i$.

To specify the prior density on the class of partitions, it therefore suffices to specify a neighborhood system (and by implication an associated clique system) and potential functions. For our purposes, the neighborhood system is defined as the entire graph, so that every region identifier is in the neighborhood of every other region identifier. This neighborhood system would generally make the implied distribution computationally intractable since the conditional distribution of any site could well depend on the values of all other sites. However, the nonzero potential functions employed in this model are easily computed and the large neighborhood system in fact poses little computational difficulty.

With these definitions, the Gibbs distribution on the region identifiers may be specified. In doing so, prior notions regarding likely configurations of objects in the true scene can be modeled, and three such properties are modeled here. First, configurations having large numbers of regions are discouraged. In noisy images, this type of constraint is needed to prevent individual pixels from assuming completely arbitrary
values, a problem that makes maximum likelihood estimation unattractive in many settings (e.g. Vardi, Shepp, and Kaufmann 1985). Second, irregular object shapes are discouraged, although the extent to which they are penalized depends on the nature of the true scene. Third, configurations containing disconnected objects are prohibited. In two-dimensional image restoration this condition may not always be sensible, although in three-dimensional image reconstruction it almost always is. Of course, the constraint can be dropped when deemed inappropriate.

The first potential type, designed to restrict the number of regions, is a function of the entire graph. The particular form of the potential depends on the nature of the image scene and the anticipated number of objects in the scene. If little prior information regarding the number of regions is available, a possible choice for this potential might be $V(R) = \alpha K$, where $K$ represents the number of distinct region identifiers in the graph and $\alpha$ is an arbitrary hyperparameter. The effect of this potential is to impose a constant penalty of $\exp(\alpha)$ on the formation of each new region. Alternatively, if prior knowledge of the image scene suggests that the number of regions should be restricted to a relatively small number, a quadratic potential of the type

$$V(R) = \alpha K^2$$

might be used.

The second class of potentials are designed to encourage regularly shaped regions. One possible potential in this category is the potential that assigns an energy of, say, $\zeta$ to each pair of neighboring pixels not assigned the same region identifier (e.g. Derin and Elliot 1987). Such potentials impose local regularity, but in some cases may not be appropriate since the increase in energy associated with long smooth boundaries is also quite large. This can cause the prior to favor configurations having a small number of regions, or a number of small regions with short boundaries.

An alternative class of regularity cliques are depicted in Figure 2 (for the remainder of the paper we consider only hexagonal lattices). Each member of this class consists of a central pixel and a ring of surrounding pixels. The diameter of the ring is set according to the diameter of likely objects in the image scene. We denote the clique type containing the central pixel and the surrounding pixels labeled $a$ by $C_a$, the cliques containing the central pixel and surrounding pixels labeled $b$ by $C_b$, and so on.

The potential for each of these cliques is determined by examining the connectivity of region identifiers within the outer ring. If for each region identifier in the ring it is possible to move to all similar region identifiers in the ring without leaving the ring, the potential assigned to the clique is zero. On the other hand, if a pixel with a given region identifier is separated within the ring from other pixels belonging to the
Figure 2. The configuration depicted (a) does not incur a regularization penalty since the voxels labeled 2 are connected even when the center pixel is removed from the graph. In contrast, the configuration depicted in (b) results in a penalty of $\phi$ since the voxels labeled 2 would no longer be connected if the center voxel was removed. In both illustrations, all voxels not outlined are assumed to be assigned with region identifier 1.

The same region, the potential assigned to the clique is $\phi$.

The final potential type is included to prevent a region from splitting into two disconnected partitions. To understand the need for such a potential, suppose that a partitioning set has the shape of a dumbbell, and that the connecting "bar" is one pixel wide. When updating any of the pixels in the bar, a change in the value of the given region identifier would separate the dumbbell into two distinct regions, requiring that all region identifiers in one of the two segments be changed. However, such changes violate the Markovian property of the Gibbs distribution, and so pixels in the bar are not permitted to change. This is accomplished by assigning infinite potentials to changes in region identifiers that result in disconnected regions. Like the constraint on the number of regions, the clique associated with this potential is the entire graph.

2.2 The Intensity Model

Having specified the prior distribution on the region identifiers, the next stage in the model associates with each partitioning set a mean intensity parameter. Individual pixel intensities within regions are assumed to be drawn from a distribution centered around this value. Typically, the form of this distribution is taken
to be conjugate to the distribution used to model the observed pixel values in the final stage of the model. In the final stage of the model, an exponential family distribution is assumed for the generation of observations at individual pixels.

As an example, consider a Poisson-gamma model for the observations and pixel intensities. In this case, let \( \mu_k \) denote the mean intensity for pixels in partitioning set \( k \), let \( \Lambda = \{ \lambda_i \} \) denote the array of pixel intensities, and let \( Y = \{ y_i \} \) denote the array of observed Poisson counts at individual pixels. Then given the region identifiers, the hierarchical model for image generation can be expressed

\[ y_i \mid \lambda_i \sim \text{Poisson}(\lambda_i), \]

\[ \lambda_i \mid r_i = k, \mu_k, \nu \sim G(\mu_k, \nu). \]

Here \( G(\mu, \nu) \) denotes a random variable with density function

\[ g(\lambda; \mu, \nu) = \frac{1}{\Gamma(\nu)} \left( \frac{\nu \lambda}{\mu} \right)^\nu \exp\{-\nu \lambda / \mu\}. \]

For the Poisson-gamma model, these conditional densities result in a joint posterior density for the parameters \( \Lambda, R, \mu, \) and \( \nu \) proportional to

\[ \exp\left\{ -\alpha K^2 + \sum_{\text{all pixels } i} \left[ (y_i + \nu - 1) \log(\lambda_i) - \lambda_i (1 + \nu / \mu_i) - \nu \log(\mu_i) + \nu \log(\nu) - \log(\Gamma(\nu)) - V_i \right] \right\}. \]

In this expression, \( K \) represents the number of distinct region identifiers in the configuration, and \( V_i \) denotes the potential associated with the specific regularity clique(s) used in the model (e.g., \( C_a \)). It is assumed that the region identifier array \( R \) is connected; otherwise the posterior density is 0.

3. DETERMINISTIC APPROACH TO THE SAMPLING DISTRIBUTION

As described in Johnson(1991c), posterior distributions for region volumes and region means can be obtained by Gibbs sampling technique. But doing so is very computational demanding. So developing some good deterministic approach to this problem is practically important and theoretically interesting. Our goal is to approximate sampling distribution of some important image features, say, region volume, region mean and maybe pixel intensity also.

Let \( r_{i1}, r_{i2} \) be the two most probable region identifiers for pixel \( \xi_i \). They may be the same or different. Let \( R = \{ r_i \} \) be the true region identifier array. Let \( p_i \) denote the probability for pixel \( \xi_i \) to be assigned the first region identifier \( r_{i1} \), and hence \( 1 - p_i \) is the probability for pixel \( \xi_i \) to be assigned the second region identifier \( r_{i2} \). By "first" we mean that \( p_i > 1 - p_i \) or \( p_i > 0.5 \). Let \( P = \{ p_i, r_{i1}, r_{i2} \} \) be the array to store
the highest probability and the two most probable region identifiers. Suppose we are particularly interested in region $k$. Let

$$J_k = \{\xi_i : r_{i1} = k \ or \ r_{i2} = k\}. \quad (3.1)$$

$$p^k_i = p_i I(r_{i1} = k) + (1 - p_i) I(r_{i2} = k)$$

$$Vol(k) = \text{volume of region } k$$

$$N = \#'s \ of \ pixels \ in \ J_k$$

Where $I(\cdot)$ is an indicator function, if $(\cdot)$ holds then $I(\cdot) = 1$, otherwise 0. If we simply treat the region identifiers as independent Bernoulli random variables--- so that the region identifier at one pixel is independent of the region identifier at any other pixels (this assumption will be adjusted later), then we have:

$$E[Vol(k)|A, P] = \sum_{\xi_i \in J_k} p^k_i. \quad (3.2)$$

$$Var[Vol(k)|A, P] = \sum_{\xi_i \in J_k} p^k_i (1 - p^k_i). \quad (3.3)$$

So given the two region identifiers and corresponding probabilities the expected volume of region $k$ and the uncertainty about it can be calculated.

Let $\{x_i = 1\}$ mean that $r_{i1} = k$, or $r_{i2} = k$, and note that $p^k_i$ is then the probability that $\{x_i = 1\}$. Considering the segmentation shown in Figure 3, by likelihood and prior only, pixels $x_1$ and $x_6$ have probabilities 0.6 and 0.8 of being in region $k$ respectively. And pixels $x_4$ and $x_6$ and all the pixels unlabelled below the dotted line in the picture have no possibility to be in region $k$. Pixels $x_1$, $x_2$, and $x_3$ and all the pixels unlabelled above the dotted line have probability 1 to be in region $k$. Since regions must be connected, the probability for $x_6$ being in region $k$ should be less than the probability for $x_0$ being in region $k$, even though $x_6$ has the greater likelihood to be in region $k$. To obtained this condition, we apply following modification to the updating procedure. First, the probability for none of $x_0$’s neighboring pixels to be in region $k$ is

$$(1 - p^k_1)(1 - p^k_2)(1 - p^k_3)(1 - p^k_4)(1 - p^k_5)(1 - p^k_6)$$

Subtracting this product from unity gives the probability that at least one of the neighbors of pixel $x_0$ is in region $k$. Let $p$ be the likelihood that $x_0$ is in region $k$. Then the updated probability for $x_0$ to be in region $k$ is:

$$p_0^k = p \ast [1.0 - (1 - p^k_1)(1 - p^k_2)(1 - p^k_3)(1 - p^k_4)(1 - p^k_5)(1 - p^k_6)]$$

7
The dependence of a pixel's region identifier on those of its neighbors implies covariances between these region identifiers. Note that $p_i^k$ is the mean of $x_i$, and so by definition of the covariance:

$$\text{Cov}(x_0, x_1) = E\{(x_0^k - p_0^k) \cdot (x_1^k - p_1^k)\}$$

$$= \sum_{i=0,1} (i - p_0^k) \cdot (j - p_1^k) \cdot p(x_0 = i, x_1 = j)$$

(3.4)

The joint probability in (3.4) can be written as:

$$p(x_0 = 0, x_1 = 0) = p(x_0 = 0|x_1 = 0) \cdot p(x_1 = 0)$$

$$p(x_0 = 1, x_1 = 0) = p(x_0 = 1|x_1 = 0) \cdot p(x_1 = 0)$$

$$p(x_0 = 0, x_1 = 1) = p(x_0 = 0|x_1 = 1) \cdot p(x_1 = 1)$$

$$p(x_0 = 1, x_1 = 1) = p(x_0 = 1|x_1 = 1) \cdot p(x_1 = 1)$$

Given that $x_1$ is not in region $k$, the probability of $x_0$ in region $k$ should be smaller than the unconditional probability $p(x_0 = 1)$. And vice versa. Say:

$$p(x_0 = 1|x_1 = 0) \leq p(x_0 = 1)$$

$$p(x_0 = 1|x_1 = 1) \geq p(x_0 = 1)$$
By this argument and the way we calculated the unconditional probability \( p(x_0 = 0) \) we get:

\[
p(x_0 = 1|x_1 = 0) = p \cdot \left[ 1 - (1 - p_2^k)(1 - p_3^k)(1 - p_4^k)(1 - p_6^k)(1 - p_0^k) \right]
\]

\[
= p_0^* \leq p_0^k
\]

which means given one of \( x_0 \)'s neighbors \( x_1 \) not being in region \( k \), the probability for \( x_0 \) being in region \( k \) becomes smaller.

\[
p(x_0 = 1|x_1 = 1) = \frac{p_0^k}{\left[ 1 - (1 - p_1^k)(1 - p_2^k)(1 - p_3^k)(1 - p_4^k)(1 - p_5^k)(1 - p_6^k) \right]}
\]

\[
= p_0^{**} \geq p_0^k
\]

which means given one of \( x_0 \)'s neighbors \( x_1 \) being in region \( k \), the probability for \( x_0 \) being in region \( k \) becomes larger. And finally by equation (3.4), we obtain following:

\[
Cov(x_0, x_1) = p_1^k(1 - p_1^k)(p_0^{**} - p_0^*). \tag{3.5}
\]

Thus variance of region volume can be modified as following by using (3.5):

\[
Var[Vol(k)|\Lambda, P] = \sum_{i=1}^{nsrc} Cov(x_i, x_j)
\]

\[
= \sum_{i \in J_k} p_i^k (1 - p_i^k) + \sum_{i,j=1, nsrc} Cov(x_i, x_j)
\]

where \( nsrc \) is the total numbers of pixels in the image.

The posterior distribution of the mean intensity of regions is (Johnson(1991c)):

\[
p(\mu_k|\Lambda, R) \propto \exp \left( - \frac{\sum_{i=1}^{r_i=k} \nu \lambda_i}{\mu_k} - \sum_{r_i=k} \nu \log(\mu_i) \right)
\]

\[
= IG \left( \sum_{r_i=k} \nu - 1, \nu \sum_{r_i=k} \lambda_i \right)
\]

Where \( IG(\cdot) \) is the inverse gamma density function. So by Bayes' theorem, given the probability and top-two-region identifiers array \( P(.) = (p_1, r_{11}, r_{12}) \) and pixel intensity-array \( \Lambda \), the distribution of \( \mu_k \) will be:

\[
p(\mu_k|\Lambda, P) \propto \int_R p(\mu_k|\Lambda, R)p(R|P)dR
\]

\[
= \int_R IG \left( \sum_{r_i=k} \nu - 1, \nu \sum_{r_i=k} \lambda_i \right) \prod_{\xi \in J_k} p_k^i (I_{r_1=k}^i) (1 - p_i) I_{r_1=\alpha} dR. \tag{3.6}
\]

\[
= \sum_{\text{2 terms}} c_k IG(\alpha_k - 1, \beta_k)
\]

where

\[
\alpha_k = \nu \left( \sum_{\xi \in J_k, \rho_r = 1} 1 + \sum_{\xi, \nu_r \neq 1} I(p_i = p_i^k) \right). \tag{3.7}
\]
\[ \beta_k = \nu \left( \sum_{\xi \in J_k, \pi_i = 1} \lambda_i + \sum_{\xi \in J_k, \pi_i \neq 1} \lambda_i I(p_i = p_i^k) \right). \quad (3.8) \]

\[ c_k = \prod_{\xi \in J_k} p_i'. \quad (3.9) \]

where \( p_i' = p_i^k \), or \( 1 - p_i^k \) and \( \sum c_k = 1 \). Then if \( K = 8 \), we'll get a mixture of \( 2^K = 256 \) different inverse gamma distributions for the \( \mu_k | \Lambda, P \). In fact, many of the \( c'_k \)s are near zero. Then we have

\[
p(\mu_k | \Lambda, P) = \sum_{k=1}^{2^K} c_k IG(\alpha_k - 1, \beta_k) \\
E'(\mu_k | \Lambda, P) = \sum_{k=1}^{2^K} c_k \frac{\beta_k}{\alpha_k - 2} \\
Var(\mu_k | \Lambda, P) = \sum_{k=1}^{2^K} c_k \left( \frac{\beta_k^2}{(\alpha_k - 2)(\alpha_k - 3)} - E(\mu_k | \Lambda, P)^2 \right)
\]

Thus at each iteration, conditional on the two most probable region identifiers and probability the expected region means can be calculated and corresponding uncertainty can be assessed from above formulae.

4. Example

To illustrate the performance of our approach, we first consider the application of our approximation method to a standard phantom used in nuclear medicine, the 2-D Hoffman brain phantom (Hoffman et al 1990). We restrict our particular interest to the region indicated by the arrow in Figure 4. For comparison, 5000 standard Gibbs sampling iterations was employed to obtain the true posterior distribution of region mean and its volume of the region indicated in Figure 4, the histograms for these posterior distributions of the area and mean of the region are shown in Figure 5(a) and (b). The true area of this region is 140 pixels and the mean intensity is 20.

The superimposed curves( dotted lines ) in Figure 5(a) and (b) show the approximate posterior distributions obtained at iteration 17 of the method presented in Section 3. The approximate posterior distributions are assumed to be normals with means and variances determined as discussed in Section 3.

Figure 6 shows us the image from our algorithm by taking the two most probable region identifier and Figure 7 shows us the image by taking only the most probable region identifier.

For the further application, we'll implement our technique to the 3-D image reconstruction. In 3-D image reconstruction, we have already implemented updating scheme for identifiers from projection data. For better initial images, we are looking at metz-filtered-back-projection method and ML method. For better initial segmentations, we are investigating Marr-Hildreth edge detection method. Hopefully by incorporating
these into our approach to the posterior distribution of image parameters of particular interest, we will get as good approximation in 3-D image reconstruction as in 2-D Hoffman case.

Figure 4. (a) The true pixel intensities of the 2-D Hoffman phantom.

Figure 4. (b) A Poisson observation of the 2-D Hoffman phantom.
Figure 5. (a) Histogram from Gibbs sampling v.s. Approximation of posterior from our method for region volume.

Figure 5. (b) Histogram from Gibbs sampling v.s. Approximation of posterior from our method for region mean intensity.
Figure 6. The reconstructed image of the phantom using the two most probable region identifiers.

Figure 7. The reconstructed image of the phantom using only the most probable region identifier.

REFERENCES


