Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 89 (2004) 338–350
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

# Wavelet modeling of priors on triangles

Dipak K. Dey* and Yazhen Wang

*Department of Statistics, University of Connecticut, U-4120, Storrs, CT 06269, USA*

Received 26 February 2001

**Abstract**

Parameters in statistical problems often live in a geometry of certain shape. For example, count probabilities in a multinomial distribution belong to a simplex. For these problems, Bayesian analysis needs to model priors satisfying certain constraints imposed by the geometry. This paper investigates modeling of priors on triangles by use of wavelets constructed specifically for triangles. Theoretical analysis and numerical simulations show that our modeling is flexible and is superior to the commonly used Dirichlet prior.


*AMS 2000 subject classifications:* 62G07; 62A15

*Keywords:* Bayesian analysis; Dirichlet distribution; Posterior distribution; Prior distribution; Simplex

## 1. Introduction

Unknown parameters in practical statistical problems often have some constraints. These constraints force the parameter space to satisfy certain geometry, and statistical inference must take the geometry into consideration. For example, suppose a map is made up of many regions, each painted in one of three colors, and we wish to estimate the proportions, $\theta_1, \theta_2, \theta_3$ of the map covered by each color. If we had a method of choosing independently and uniformly at random, $n$ points on the map and it gave $n_i$ points in color labeled by index $i$, then $n_1, n_2, n_3$ have the trinomial distribution, where $n = n_1 + n_2 + n_3$. Clearly, the parameter space $T = \{\theta = (\theta_1, \theta_2, \theta_3)\colon\ 0 \leqslant \theta_i \leqslant 1, \sum_{i=1}^3\ \theta_i = 1\}$ is a simplex. In order to perform a Bayesian analysis, prior distributions for $\theta$ must be selected to live on the corresponding

*Corresponding author. Fax: +1-860-486-4113.
*E-mail address:* dey@stat.uconn.edu (D.K. Dey).

0047-259X/03/$ - see front matter 
doi:10.1016/S0047-259X(03)00080-0

geometry, i.e., simplex in this case. Here we consider the case that parameters belong to a triangle. Dirichlet priors are commonly used in the literature when the parameter space is a simplex. For example, West [7] used Dirichlet prior for modeling probabilities for expert opinion. As an alternative to Dirichlet prior, Aitichison [1] developed a general class of distributions on the simplex. Gelfand et al. [4] used a mixture of Beta prior, Wolpert and Lavine [8] used Markov random field, and Iyengar and Dey [5] used generalized Liouville distribution when the parameter space is restricted to a simplex. In this paper, we employ wavelets which are specifically constructed to model priors on the triangle. The wavelet-based priors are not only conceptually simple and very flexible but also can well approximate all practically reasonable prior distributions. Moreover, the properties of wavelets allow us to use available information to specify priors with certain features.

The rest of the paper is organized as follows. Section 2 develops multiwavelets on a triangle. In Section 3, we show that the wavelet-based priors are flexible and can well approximate any true priors on the triangle. We further show that the posterior distributions and Bayesian estimators resulted from using the wavelet-based prior and the true prior are very close. Thus, for prior elicitation in a Bayesian analysis, a wavelet-based prior is quite robust. Finally, results based on simulations are developed in Section 4 to demonstrate effectiveness of the wavelet-based prior and compare with Dirichlet priors under various scenarios.

## 2. Multiwavelets on a triangle

Wavelets are typically constructed on simple domains like Euclidian spaces and their rectangular subsets. To construct wavelets on a triangle we split the triangle into four same-size small triangle by midpoint subdivision and continue the splitting again and again to obtain self-similar triangles. The construction of wavelets on the square makes quite extensive use of Fourier analysis. However, on complicated domains like triangles, Fourier analysis becomes less accessible, so simple polynomial operations are used to design wavelets on these domains.

### 2.1. Multiresolution analysis on triangles

Let $T$ be a triangle. Consider its successive refinement $\{T_{j,k};\ j\geqslant 1, k\in\mathscr{I}^j\}$, where $\mathscr{I}^j = \{1, \dots, 4^j\}$, and each triangle in a finer scale is constructed from one in a coarser level by midpoint subdivision, denoted the resulting three subtriangles by

$$T_{j,k} = T_{j+1,k_0} \cup T_{j+1,k_1} \cup T_{j+1,k_2} \cup T_{j+1,k_3}.$$

For consistency, let $T_0 = T$. We use the convention that the center and the four peripheral subtriangles in the midpoint subdivision of $T_{j,k}$ are indexed by $T_{j+1,k_0}$, $T_{j+1,k_1}$, $T_{j+1,k_2}$, $T_{j+1,k_3}$, respectively; and when no confusion occurs, simply by $T_0, T_1, T_2, T_3$. The edges of $T_{j+1,k_0}$ are indexed by $e^1_{j,k}, e^2_{j,k}, e^3_{j,k}$ or simply by $e_1, e_2, e_3$, such that $e_i$ represents the common boundary of $T_0$ and $T_i$. For $d \geqslant 0$, let

$$P_d = \{x^i y^j; i+j \leqslant d\}, \quad P_d(T) = \{f;\ f|_T \in P_d,\ f|_{R^2 \setminus T} = 0\},$$

where $f|_T$ denotes the restriction of $f$ on $T$. Then $dim(P_d) = dim(P_d(T)) = M = (d+1)(d+2)/2$. For

$$T = T_0 \cup T_1 \cup T_2 \cup T_3,$$

define

$$V = P_d(T_0) \oplus P_d(T_1) \oplus P_d(T_2) \oplus P_d(T_3) \quad \text{and} \quad W = V \ominus P_d(T).$$

Then $dim(V) = 4M$ and $dim(W) = 3M$. Now we will construct orthogonal basis for $P_d(T)$ and $W$. We denote an orthonormal basis for $P_d(T)$ by $\{\chi_T^\ell;\ 0 \leqslant \ell < M\}$ or simply by $\chi_T$ as a vector of functions and an orthonormal basis for $W$ by $\{h_{e_i}^\ell; 0 \leqslant \ell < M,\ i = 1,2,3\}$, or simply by the vector notation $h_1, h_2, h_3$. Given such an orthonormal basis, we have orthonormal basis for $L^2(T)$:

$$\{\chi_{T_0}, h_{e^i_{j,k}};\ i = 1,2,3,\ j \geqslant 0,\ k \in \mathscr{I}^j\},$$

and any $f \in L^2(T)$ has a decomposition

$$f = \alpha_0 \chi_{T_0} + \sum_{i=1}^{3} \sum_{j=0}^{\infty} \sum_{k \in \mathscr{I}^j} \beta^i_{j,k} h_{e^i_{j,k}},$$

where

$$\alpha_0 = \langle f, \chi_{T_0} \rangle = (\langle f, \chi^0_{T_0} \rangle, \ldots, \langle f, \chi^{M-1}_{T_0} \rangle)$$

and

$$\beta^i_{j,k} = \langle f, h_{e^i_{j,k}} \rangle = (\langle f, h^0_{e^i_{j,k}} \rangle, \ldots, \langle f, h^{M-1}_{e^i_{j,k}} \rangle).$$

For a nested triangular tessellation $T_{j,k}$ of $T$, let

$$V_j = \bigoplus_{k \in \mathscr{I}^j} P_d(T_{j,k})$$

be the space of piecewise polynomials of degree less than $d$ on each of $T_{j,k}$. Then multiresolution analysis (MRA) on the triangle $T$ is given by

(1) $V_j \subset V_{j+1}$;
(2) $\lim_{j \to \infty} V_j = L^2(T)$;
(3) $\{\chi^\ell_{T_{j,k}}; 0 \leqslant \ell < M, k \in \mathscr{I}^j\}$ form an orthonormal basis for $V_j$;

(4) Each of $\chi^{\ell}_{T_{j,k}}$ is compactly supported at $T_{j,k}$;
(5) For each $T_{j,k}$, there exists four $M$ x $M$ matrices $H_i$ such that

$$\chi_{T_{j,k}} = \sum_{i=0}^{3} H_i \chi_{T_{j+1,k_i}}.$$

### 2.2. Barycentric coordinates

Barycentric coordinate is convenient to work with in the following construction of wavelets on a triangle. Let $P_i = (x_i, y_i)$ be the ordered list of three vertices of $T$. A point $P = (x, y)$ can be expressed in terms of its barycentric coordinates $\tau = (\tau_1, \tau_2, \tau_3)_T$ with respect to $T$ as follows:

$$P = \sum_{i=1}^{3} \tau_i P_i,$$

where $\sum_{i=1}^{3} \tau_i = 1$. If $P$ is inside $T$, then $\tau_i \geqslant 0$.

Change of coordinates from Cartesian to barycentric can be easily carried out by the following transformation:

$$\begin{pmatrix} x(\tau_1, \tau_2) \\ y(\tau_1, \tau_2) \end{pmatrix} = \begin{pmatrix} x_1 - x_3 & x_2 - x_3 \\ y_1 - y_3 & y_2 - y_3 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} x_3 \\ y_3 \end{pmatrix}.$$

Changing barycentric coordinate from one reference triangle to another obeys the following rule. If $P$ has barycentric coordinates $\tau' = (\tau_1{}', \tau_2{}', \tau_3{}')_{T'}$ relative to another triangle $T'$ defined by three vertices $\{(x_i{}', y_i{}');\ i = 1, 2, 3\}$, then

$$(\tau_1{}', \tau_2{}', \tau_3{}')_{T'} = (\tau_1, \tau_2, \tau_3)_T M_{T \to T'},$$

where

$$M_{T \to T'}{}' = \begin{pmatrix} x_1{}' & x_2{}' & x_3{}' \\ y_1{}' & y_2{}' & y_3{}' \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}.$$

Midpoint subdivision can be easily described using barycentric coordinates by restricting $\tau_i$ bounded below or above $\frac{1}{2}$.

Thus, with a wavelet basis constructed on a triangle $T$, using the property of barycentric coordinate we easily obtain a wavelet basis on triangle $T'$.

### 2.3. Construction of scaling functions

Scaling functions $\chi^{\ell}_{T_0}$ are orthogonal polynomials supported by $T_0$ so that

$$\langle \chi^{\ell}_{T_0}, \chi^{\ell'}_{T_0} \rangle = \delta_{\ell\ell'}.$$

The construction is based on Legendre-type polynomial basis. Due to the mutilation property, we first construct the basis on triangle $T_B$ with vertices $(0,0),(0,1),(1,0)$. Since the power basis $\{1,x,y,x^2,xy,y^2,\dots,\}$ is equal to $\{1,\tau_2,1-\tau_1-\tau_2,\tau_2^2,(1-\tau_1-\tau_2)\tau_2,(1-\tau_1-\tau_2)^2,\dots,\}$ in barycentric coordinates with respect to $T_B$, applying the Gram–Schmidt process to the latter sequence we get Legendre polynomials $\{\pi^\ell;\ell\geqslant 0\}$ on $T_B$.

Let $\chi^\ell_{T_B}=\pi^\ell 1_{T_B}$. Then the resulting sequence $S=\{\chi^\ell_{T_B}\}$ will be a triangular sequence of orthogonal polynomials: for any $d\geqslant 0$, the first $M=(d+1)(d+2)/2$ elements of $S$ form an orthonormal basis for $P_d(T_B)$. Here are the first few members of $S$:

$$\chi^0_{T_B}(\tau_1,\tau_2)=\sqrt{2}1_{T_B}(\tau_1,\tau_2),$$
$$\chi^1_{T_B}(\tau_1,\tau_2)=(-2+6\tau_2)1_{T_B}(\tau_1,\tau_2),$$
$$\chi^2_{T_B}(\tau_1,\tau_2)=2\sqrt{3}(1-2\tau_1-\tau_2)1_{T_B}(\tau_1,\tau_2),$$
$$\chi^3_{T_B}(\tau_1,\tau_2)=\sqrt{6}(1-8\tau_2+10\tau_2^2)1_{T_B}(\tau_1,\tau_2),$$
$$\chi^4_{T_B}(\tau_1,\tau_2)=3\sqrt{2}(-1+2\tau_1+6\tau_2-10\tau_1\tau_2-5\tau_2^2)1_{T_B}(\tau_1,\tau_2),$$
$$\chi^5_{T_B}(\tau_1,\tau_2)=\sqrt{30}(1-6\tau_1+6\tau_1^2-2\tau_2+6\tau_1\tau_2+\tau_2^2)1_{T_B}(\tau_1,\tau_2).$$

Mutilation gives us the scaling function as

$$\chi^\ell_{T_{j,k}}=\sqrt{\frac{1}{2|T_{j,k}|}}\chi^\ell_{T_B},\quad 0\leqslant\ell<M.$$

### 2.4. Mother multiwavelets

With an orthonormal basis for $P_d(T)$, we now construct an orthonormal basis for

$$W=\{P_d(T_0)\oplus P_d(T_1)\oplus P_d(T_2)\oplus P_d(T_3)\}\ominus P_d(T).$$

Let $\{\pi^\ell_T\}$ be the Legendre polynomials mutilated to the triangle $T$. Define for $i=1,2,3,\ \ell=0,1,\dots,M-1$,

$$h^\ell_i(\tau_1,\tau_2,\tau_3)=\begin{cases}\pi^\ell_{T_i}((\tau_1,\tau_2)M_{T\to T'}) & \text{on } T_i,\\ -\pi^\ell_{T_i}((\tau_1,\tau_2)M_{T\to T'}) & \text{on } T\backslash T_i,\\ 0 & \text{otherwise.}\end{cases}$$

Note that $\{h^\ell_i\}\cup\{\chi^\ell_T\}$ spans $P_d(T_0)\oplus P_d(T_1)\oplus P_d(T_2)\oplus P_d(T_3)$. We will modify successively to make an orthonormal basis for $W$. First orthogonalize $\{h^\ell_i\}$ against $P_d(T)$ and replace $\{h^\ell_i\}$ by

$$h^\ell_i-\sum_{\ell=0}^{M-1}\langle h^\ell_i,\chi^\ell_T\rangle\chi^\ell_T,$$

and denote the above functions still by $\{h_i^\ell\}$. Then $\text{span}(\{h_i^\ell\}) = P_d(T)^\perp \cap \oplus_{i=1}^3 P_d(T_i)$. Orthogonalizing $\text{span}\{h_i^\ell\}$ by Gram–Schmidt, we get an orthonormal basis for $W$. Once again denote the basis by $\{h_i^\ell\}$. Finally, define the mother multiwavelets as

$$h_{e_{j,k}^i}^\ell = \sqrt{\frac{1}{2|T_{j,k}|}}\, h_i^\ell, \quad 0 \leqslant \ell < M.$$

Note that, multiwavelets with better smoothness, more symmetry and/or vanishing moments can be constructed by refining the design scheme.

## 3. Prior approximation

Suppose a prior $\pi$ for a parameter $\theta$ belonging to a triangle $T$ has a probability density function $\pi(\theta)$. Then we have the wavelet expansion for $\pi(\theta)$ as

$$\pi(\theta) = \alpha_0 \chi_{T_0}(\theta) + \sum_{i=1}^{3} \sum_{j=0}^{\infty} \sum_{k \in \mathscr{I}^j} \beta_{j,k}^i h_{e_{j,k}^i}(\theta). \tag{1}$$

Denoted by $\Pi$ the class of priors whose densities have above wavelet expansion.

Given a prior distribution $\pi(\theta)$, for any $\varepsilon > 0$, there is an approximation $\pi_\varepsilon(\theta) \in \Pi$ with wavelet expansion up to $J$ levels:

$$\pi_\varepsilon(\theta) = \alpha_0 \chi_{T_0}(\theta) + \sum_{i=1}^{3} \sum_{j=0}^{\infty} \sum_{k \in \mathscr{I}^j} \beta_{j,k}^i h_{e_{j,k}^i}(\theta) \tag{2}$$

such that

$$||\pi_\varepsilon(\theta) - \pi(\theta)|| \leqslant \varepsilon, \tag{3}$$

where $|| \cdot ||$ denotes the total variation norm. Suppose $X_1, \ldots, X_n$ are i.i.d. with probability density function $f(x|\theta)$. The posterior probability density function is then given by

$$\pi(\theta|\mathbf{X}_n) = \pi(\theta) f(\mathbf{X}_n|\theta)/f(\mathbf{X}_n),$$

where $\mathbf{X}_n = (X_1, \ldots, X_n)$, $f(\mathbf{X}_n|\theta) = \prod_{i=1}^n f(X_i|\theta)$, and $f(\mathbf{X}_n) = \int f(\mathbf{X}_n|\theta)\pi(d\theta)$ is the marginal probability density function.

Denote by $f_\varepsilon(\mathbf{X}_n)$ and $\pi_\varepsilon(\theta|\mathbf{X}_n)$ the respective marginal probability density function of $\mathbf{X}_n$ and the posterior probability density function of $\theta$ given $\mathbf{X}_n$, when the prior probability density function is taken to be $\pi_\varepsilon(\theta)$. Now we have the following theorem.

**Theorem 1.** *Assume $f(x|\theta)$, as a function of $\theta$, is bounded from above for each $x$, then*

$$||\pi_\varepsilon(\theta|\mathbf{X}_n) - \pi(\theta|\mathbf{X}_n)|| \leqslant K(\mathbf{X}_n)\varepsilon,$$

*where $K(\mathbf{X}_n)$ depends on data $\mathbf{X}_n$ only.*

**Proof.** Denote by $K$ a generic constant whose value may change from appearance to appearance and let $v = \pi_\varepsilon - \pi$. Then by Hahn decomposition, we have $v = v^+ - v^-$, and $||v|| = v^+(T) + v^-(T)$, where $v^+$ and $v^-$ are finite measures with mutually exclusive supports. For the absolute difference of two marginal probability density functions, we have

$$
\begin{aligned}
|f_\varepsilon(\mathbf{X}_n) - f(\mathbf{X}_n)| &= \left| \int f(\mathbf{X}_n|\theta)\{\pi_\varepsilon(d\theta) - \pi(d\theta)\} \right| \\
&= \left| \int f(\mathbf{X}_n|\theta) v(d\theta) \right| \\
&= \left| \int f(\mathbf{X}_n|\theta) v^+(d\theta) - \int f(\mathbf{X}_n|\theta) v^-(d\theta) \right| \\
&\leqslant \int f(\mathbf{X}_n|\theta) v^+(d\theta) + \int f(\mathbf{X}_n|\theta) v^-(d\theta) \\
&\leqslant \max_{\theta \in T}\{f(\mathbf{X}_n|\theta)\}\{v^+(T) + v^-(T)\} \\
&= K||\pi_\varepsilon - \pi|| \\
&\leqslant K\varepsilon. 
\end{aligned} \tag{4}
$$

In view of (4), we obtain

$$
\begin{aligned}
||\pi(\theta|\mathbf{X}_n) - \pi(\theta|\mathbf{X}_n)|| &= [f_\varepsilon(\mathbf{X}_n) f(\mathbf{X}_n)]^{-1} ||f(\mathbf{X}_n|\theta)\{\pi_\varepsilon(\theta) f(\mathbf{X}_n) - \pi(\theta) f(\mathbf{X}_n) \\
&\quad + \pi(\theta) f(\mathbf{X}_n) - \pi(\theta) f_\varepsilon(\mathbf{X}_n)\}|| \\
&\leqslant [f^2(\mathbf{X}_n)(1 - K\varepsilon)]^{-1} \max_{\theta \in T}\{f(\mathbf{X}_n|\theta)\}\{f(\mathbf{X}_n)||\pi_\varepsilon(\theta) - \pi(\theta)|| \\
&\quad + ||[f(\mathbf{X}_n) - f_\varepsilon(\mathbf{X}_n)]\pi(\theta)||\} \\
&\leqslant [f^2(\mathbf{X}_n)(1 - K\varepsilon)]^{-1} \max_{\theta \in T}\{f(\mathbf{X}_n|\theta)\}\{f(\mathbf{X}_n)||\pi_\varepsilon(\theta) - \pi(\theta)|| \\
&\quad + |f(\mathbf{X}_n) - f_\varepsilon(\mathbf{X}_n)|\} \\
&\leqslant [f^2(\mathbf{X}_n)(1 - K\varepsilon)]^{-1} \max_{\theta \in T}\{f(\mathbf{X}_n|\theta)\}\{f(\mathbf{X}_n)\varepsilon + K\varepsilon\} \\
&= K(\mathbf{X}_n)\varepsilon.
\end{aligned}
$$

**Corollary 1.**

$$||E(\theta|\mathbf{X}_n) - E_\varepsilon(\theta|\mathbf{X}_n)||_2 \leqslant K_1(X_n)\varepsilon,$$

*where* $||\cdot||_2$ *is the Euclidian distance.*

**Proof.** Let $\theta = (\theta_1, \theta_2, \theta_3)$. For each $\theta_i$, using the similar argument for proving (4) we show

$$\begin{aligned}|E(\theta_i|\mathbf{X}_n) - E_\varepsilon(\theta_i|\mathbf{X}_n)| &= \left|\int \theta_i \pi_\varepsilon(d\theta|\mathbf{X}_n) - \int \theta_i \pi(d\theta|\mathbf{X}_n)\right| \\ &\leqslant \max_{\theta\in T}\{|\theta_i|\}||\pi(\theta|\mathbf{X}_n) - \pi(\theta|\mathbf{X}_n)|| \\ &\leqslant \max_{\theta\in T}\{|\theta_i|\}K(\mathbf{X}_n)\varepsilon.\end{aligned}$$

We complete the proof by applying

$$||E(\theta|\mathbf{X}_n) - E_\varepsilon(\theta|\mathbf{X}_n)||_2^2 = \sum_{i=1}^{3} |E(\theta_i|\mathbf{X}_n) - E_\varepsilon(\theta_i|\mathbf{X}_n)|^2.$$

Note that the wavelet expansion (1) requires very mild condition like square integrability. In fact, all practically reasonable prior densities on a triangle satisfy it. Thus, for any prior on the triangle, we can have a wavelet-based prior (2) on the triangle which is very close to the true prior in total variation distance as given in (3). Theorem 1 and its corollary imply that, as the wavelet-based priors well approximate the true prior, the corresponding posterior distributions and the posterior means are very close to each other. Thus, the wavelet-based prior is quite robust for Bayesian analysis.

Wavelet-based prior can be used for either approximating an entire prior distribution or specifying a prior with certain features. In practice, we often have some knowledge on the prior for a given problem. For example, with enough prior information we may choose a known distribution as the prior; or we may have some feature information on the prior such as where and/or how the prior concentrates. For modeling the prior living on a triangle, wavelet-based prior can conveniently incorporate such prior knowledge into prior specification. For example, if a complicated distribution is selected as the prior, then we can use wavelet-based prior approximation (2) with the coefficients

$$\beta^i_{j,k} = \int h_{e^i_{j,k}}(\theta)\pi(\theta)\theta, \tag{5}$$

evaluated analytically or numerically from the prior $\pi(\theta)$. Alternatively, available information may enable us to envision that the prior assigns most of its mass around certain areas or points. The wavelet-based prior specification can easily accommodate such prior feature. Consider triangle $T = \{\theta = (\theta_1, \theta_2, \theta_3)\colon 0 \leqslant \theta_i \leqslant 1$ and $\sum_{i=1}^3 \theta_i = 1\}$. Suppose the prior is concentrated around two points $(\theta_1, \theta_2, \theta_3) = (\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$ and $(\theta_1, \theta_2, \theta_3) = (\frac{1}{8}, \frac{3}{4}, \frac{1}{8})$. Then we know that the important wavelets $(h_{e^1_{j,k}}, h_{e^2_{j,k}}, h_{e^3_{j,k}})$ in (2) are those whose associated triangles $(e^1_{j,k}, e^2_{j,k}, e^3_{j,k})$ contain at least one of these two points. For simplicity, we take $J = 2$ in (2), i.e., wavelet approximation up to level two. There are total 21 triangles (including $T$), with 4 in level one and 16 in level two. Each point is inside one of the two triangles in level one that are near the two original acute angles. In level two, each point sits on the common edge of two adjacent triangles. So out of the 21 triangles, seven triangles ($T$, two in level one and four in level two) contain at least one point. Thus, we select the seven corresponding wavelets in (2) and determine the wavelet coefficients $\beta^i_{j,k}$ according to (5) with available information on the prior $\pi(\theta)$. For example, without further information we may treat wavelets in the same level equally and specify their coefficients $\beta^i_{j,k}$ as in the case of wavelets on an interval (see [2,3,6]).

## 4. Simulations

In this section, we conduct a Monte Carlo simulation study to illustrate the use of wavelet-based priors in a Bayesian analysis. In the simulation study, we consider $f(x|\theta)$ is a trinomial distribution. We take $X = (X_1, X_2, X_3)$ which follows a trinomial distribution with $n = 15$ and $\theta \in T$ with $T = \{\theta = (\theta_1, \theta_2, \theta_3): 0 \leqslant \theta_i \leqslant 1 \text{ and } \sum_{i=1}^3 \theta_i = 1\}$. We simulate three prior scenarios. These priors are constructed as follows. As a probability density function on $T$, $\pi(\theta)$ is proportional to the truncation on $T$ of the distribution

$$p_1 N(\frac{1}{4}, 0.1^2) * N(\frac{1}{4}, 0.1^2) + p_2 N(\frac{3}{4}, 0.1^2) * N(\frac{1}{8}, 0.1^2) + p_3 N(\frac{1}{8}, 0.1^2) * N(\frac{3}{4}, 0.1^2),$$

which is the mixture of three bivariate normal distributions: product of two independent $N(\frac{1}{4}, 0.1^2)$, product of independent $N(\frac{1}{8}, 0.1^2)$ and $N(\frac{3}{4}, 0.1^2)$, and product of independent $N(\frac{3}{4}, 0.1^2)$ and $N(\frac{1}{8}, 0.1^2)$, with corresponding mixture weights $p_1, p_2$ and $p_3$, i.e., $\sum_{i=1}^3 p_i = 1$ and $0 < p_i < 1$ for $i = 1, 2, 3$. We approximate this prior by its wavelet expansion (1) up to the level $J = 12$.

For this prior and its wavelet approximation, there are no closed-form expansions for the posterior distributions. In view of that we calculate posterior distributions, posterior means and Bayes risks by Monte Carlo method. We simulate a sample of $\theta$ from each prior distribution and for each simulated value of $\theta$, a value of $X$ is generated from a trinomial distribution with $k = 15$, i.e., $X$ takes value in $\{0, \dots, 15\}^3$. This simulation procedure is repeated 20,000 times to obtain realizations of $(\theta, X)$. We group the 20,000 data into subsamples according to the values of $X$. For each value of $X$, we select its corresponding subsample and take the sample distribution and sample mean as the posterior distribution and posterior mean, respectively. With all posterior means, we finally calculate the sample average of the squared differences between $\theta$ and its posterior mean as Bayes risk.

In the first example, $p_1 = \frac{1}{2}$ and $p_2 = p_3 = \frac{1}{4}$. The prior is displayed in Fig. 1, and it has three modes with the largest one in the middle. The difference between the true prior and its wavelet expansion is invisible. The total variation between the true prior and the wavelet approximation is computed to be 0.005. The Bayes risks for the true prior and the wavelet approximation both are computed to be 0.0236. As a comparison, we also select a Dirichlet prior $D(7.4, 7.4, 6.7)$ to approximate the prior by matching the first two moments, and obtain the Bayes risk under the Dirichlet prior as 0.0628. The Bayes risk under the Dirichlet prior is 2.65 times of that for the wavelet prior.

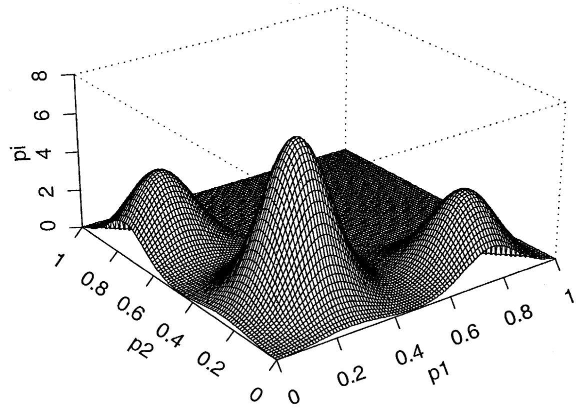



Fig. 1. The three-mode prior with the largest mode in the middle.

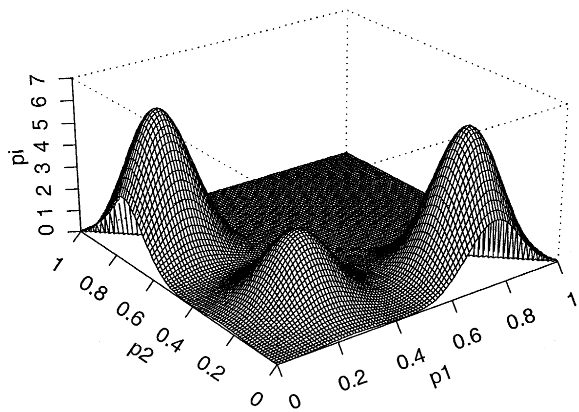


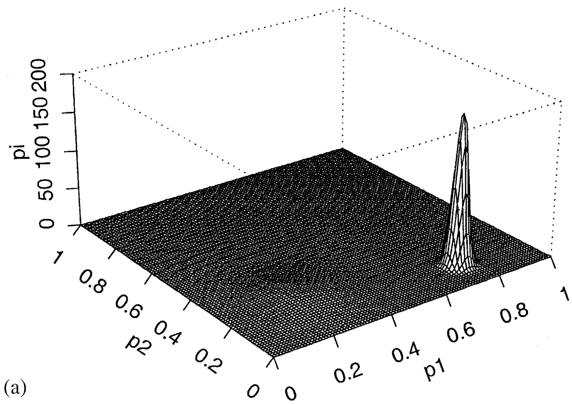

Fig. 2. The three-mode prior with the smallest mode in the middle.

In the second example, $p_1 = \frac{1}{5}$ and $p_2 = p_3 = \frac{2}{5}$. The prior has two larger modes near the triangular corners and the smallest one in the middle and is displayed in Fig. 2. From the graph, it is clear that the wavelet prior has invisible difference from the true prior. It is harder for Dirichlet prior to approximate a prior of this kind of shape. Indeed, the Bayes risk under the Dirichlet approximation prior is 3.8 times of that under the wavelet prior.

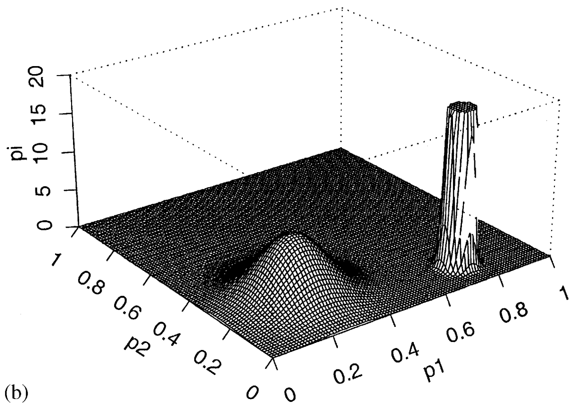
The third example is to model a prior information where $\theta$ comes from two sources which live in an area centered at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and is concentrated at a point $(\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$. Thus, we take $\pi(\theta)$ to be proportional to the truncation on $T$ of the distribution

$$\frac{1}{2}N(\frac{1}{4}, 0.1^2) * N(\frac{1}{4}, 0.1^2) + \frac{1}{2}N(\frac{3}{4}, 0.02^2) * N(\frac{1}{8}, 0.02^2)$$

which is a half and half mixture of the product of two independent $N(\frac{1}{4}, 0.1^2)$ and the product of $N(\frac{3}{4}, 0.02^2)$ and $N(\frac{1}{8}, 0.02^2)$. This corresponds to the case that the prior




(a)




(b)

Fig. 3. (a) The prior with a mode and a sharp spike. (b) The lower part of (a).

has a very sharp spike at $\theta = (\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$. The difference between the true prior and its wavelet expansion is again invisible. The total variation between the true prior and the wavelet approximation is computed to be 0.06. The ratio for the two Bayes risks for the Dirichlet approximation prior and the wavelet prior is 4.1.

From the examples we also see that modeling with wavelet-based priors retains important prior feature characteristics. For example, in Fig. 3 the histogram is displayed for wavelet-based posterior of $\theta_2$ with $X = (8, 4, 3)$. It clearly indicates that the posterior has two modes as that for the true posterior, while Dirichlet-based prior always has one mode (see Fig. 4).

## 5. Conclusion

This paper applies wavelets constructed on a triangle to model prior on the triangle in Bayesian analysis. It shows that practically reasonable prior densities can

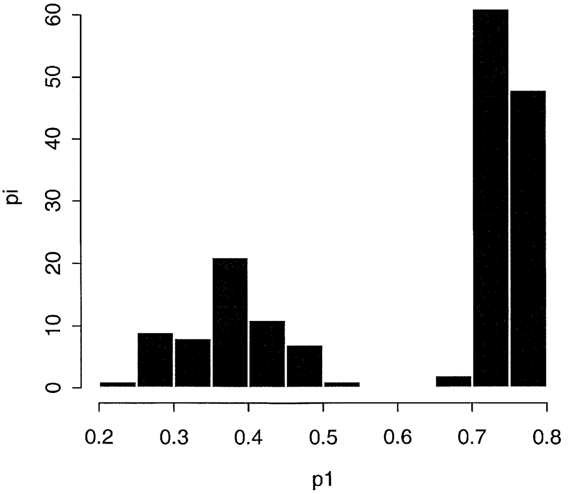



Fig. 4. The histogram for the computed posterior of $p_1$.

be well approximated by wavelet-based priors and the posterior corresponding to wavelet-based priors can be made to arbitrarily close to that based on the true prior. Modeling with wavelet-based priors is very flexible and has the ability to preserve some envisioned prior features. Numerical simulations are conducted to illustrate the effectiveness and flexibility of wavelet-based priors.

Although this paper focuses on triangles. The methodology can be easily adopted to other complicated domains like simplex and sphere. With wavelets constructed on these domains, we can easily establish the similar methodology and results for these domains.

## Acknowledgments

The authors would like to thank an anonymous referee for valuable comments and suggestions. Yazhen Wang's research is partially supported by NSA Grant MDA 904-99-1-0028 and NSF Grant DMS-0103607.

## References

[1] J. Aitchison, A general class of distributions on the simplex, J. Roy. Statist. Soc. Ser. B 47 (1985) 136–146.

[2] H.A. Chipman, E.D. Kolaczyk, R.E. McCulloch, Adaptive Bayesian wavelet shrinkage, J. Amer. Statist. Assoc. 92 (1997) 1413–1421.

[3] M. Clyde, G. Parmigiani, B. Vidakovic, Multiple shrinkage and subset selection in wavelets, Biometrika 85 (1998) 391–402.

[4] A.E. Gelfand, B. Mallick, D.K. Dey, Modeling expert opinion arising as a partial probabilistic specification, J. Amer. Statist. Assoc. 90 (1995) 598–604.
[5] M. Iyengar, D. Dey, Bayesian analysis of compositional data, in: D.K. Dey, et al., (Eds.), Generalized Linear Models: A Bayesian Perspective, Marcel Dekker, New York, 2000, pp. 349–364.
[6] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, New York, 1999.
[7] M. West, Combining expert opinion, in: J.M. Bernardo, et al., (Eds.), Bayesian Statistics, Oxford University Press, Oxford, UK, 1988, pp. 375–390.
[8] R.L. Wolpert, M. Lavine, Markov random field prior for univariate densities estimation, in: J.O. Berger, et al. (Eds.), Bayesian Robustness, IMS Lecture Notes Monograph Series, 1996, pp. 253–266.