1 Introduction and motivation

It is very common to use the distribution of pixel grey levels in an image as a synthetic description of this image. However it may be more interesting and more informative, specially for very large size images, to consider what can be called local distributions, that is distributions of grey levels within a neighbourhood of each pixel.

Considering these local distributions as given random elements, it can be of interest to describe how these local distributions are, at their turn, distributed. In other words, we would like to define a notion of distribution of local distributions. The proper mathematical tool to handle with such a description is the notion of random distribution. Indeed, as said in the nice paper of J.F.C. Kingman [14] on discrete random distributions, this notion is needed when dealing with the description of random elements which are themselves probability distribution.

A random distribution (RD) is a measurable map from a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ to the space $P(V)$ of all probability measures defined on a fixed measurable set $(V, \mathcal{V})$.

In the above example of images, the set $V$ can be taken as the range of grey levels.

Various application domains are mentioned in Pitman and Yor [19] : models in ecology, models in populations genetics, models in storage and search, zero sets of stochastic processes, asymptotic distributions in number theory, representation of partition structures and so on. An other important domain is nonparametric Bayesian statistics with the Dirichlet RD introduced in a famous paper of T.S. Ferguson [11] and further developped in C. Antoniak [3].

In this paper, we introduce RDs in image analysis by proposing new hierarchical image models and a new segmentation method. More precisely, in our model, an image is considered as being composed of distinct clusters such that the local distributions in each cluster are generated by a fixed random distribution while the whole set of local distributions is generated by a mixture of such random distributions. Further, once a local distribution is generated, this distribution generates the pixels in the corresponding neighbourhood.

It is worthwhile to note that these models can be used for clustering images just as it was done for clustering internet flows (Emilion [10], Soule et al. 2004). Finally we mention that an approach of distribution of distributions using copulas is proposed by Diday [7].
The paper is organized as follows: In Section 3 we recall the celebrated Dirichlet discrete RD of Ferguson. Then in Section 4, we study a continuous RD defined by Kraft and we prove that this RD can be considered as a randomized Haar wavelet serie. For image analysis purposes, we use an extension of Kraft construction in dimension 2. In Section 5, we propose an algorithm for estimating a finite mixture of RDs. In Section 6, we present a new hierarchical model of images and also a new segmentation algorithm based on both Dirichlet and Kraft RD and Polya urn scheme. Section 6 is followed by concluding remarks in Section 7.

2 Basic notations

Let \((\Omega, \mathcal{F}, \mathcal{P})\) be a probability space and let \(P(V)\) denote the space of all probability measures defined on a measurable set \((V, \mathcal{V})\). We denote by \(\mathcal{B}\) the smallest \(\sigma\)-field on \(P(V)\) such that all the mappings

\[
f_A : (P(V), \mathcal{B}) \to [0, 1]
\]

\[
Q \mapsto f_A(Q) = Q(A)
\]

are measurable for any \(A \in \mathcal{V}\).

**Definition 2.1** A random distribution \(X\) is a measurable mapping from \((\Omega, \mathcal{F}, \mathcal{P})\) to \((P(V), \mathcal{B})\)

\[
X : (\Omega, \mathcal{F}, \mathcal{P}) \to (P(V), \mathcal{B})
\]

\[
\omega \mapsto X(\omega)
\]

We denote by \(\mathcal{P}_X\) the distribution of \(X\) that is the probability measure on \((P(V), \mathcal{B})\) defined by:

\[
\mathcal{P}_X : \mathcal{B} \to \mathbb{R}
\]

\[
C \mapsto \mathcal{P}_X(C) = \mathcal{P}(X^{-1}(C)) = \mathcal{P}(\omega \in \Omega : X(\omega) \in C)
\]

3 Discrete random distributions

3.1 Random distributions on a finite set \(V\)

In this case, if \(l\) denotes the cardinality of \(V\), then \(P(V)\) can be identified to the set

\[
E_l = \{y = (y_1, ..., y_l), \; y_j \geq 0, \; \sum_{j=1}^{l} y_j = 1\}.
\]

Distributions on \(P(V)\) can therefore be obtained in considering the distribution of a positive random vector divided by the sum of its coordinates.

This is the case of standard Dirichlet distributions:
3.1.1 Dirichlet distribution \( D(\alpha_1, ..., \alpha_l) \)

Let \( \alpha = (\alpha_1, ..., \alpha_l) \) with \( \alpha_i > 0 \), and let \( (Z_1, ..., Z_l) \) be \( l \) independent random variables with gamma densities \( \gamma(1, \alpha_1), ..., \gamma(1, \alpha_l) \) respectively, where

\[
\gamma(b, a)(x) = \frac{1}{\Gamma(a)} b^a e^{-bx} x^{a-1} 1_{(x>0)}
\]

where \( 1_{(x>0)} \) is the characteristic function of the set \((x>0)\).

The Dirichlet distribution \( D(\alpha_1, ..., \alpha_l) \) is defined as the distribution of the random vector \((Y_1, ..., Y_l) = (Z_1 / Z, ..., Z_l / Z)\), where \( Z = Z_1 + ... + Z_l \).

This distribution is singular with respect to Lebesgue measure since its support has Lebesgue measure 0. However if \( l \geq 2 \), then \((Y_1, ..., Y_{l-1})\) has the following density:

\[
d(\alpha|y) = \frac{\Gamma(\alpha_1 + ... + \alpha_l)}{\Gamma(\alpha_1)...\Gamma(\alpha_l)} y_1^{\alpha_1-1} ... y_{l-1}^{\alpha_{l-1}-1}(1 - \sum_{h=1}^{l-1} y_h)^{\alpha_l-1} S_l(y),
\]

where \( S_l = \{y = (y_1, ..., y_{l-1}), y_j \geq 0, \sum_{j=1}^{l-1} y_j < 1\}\).

It is more convenient in concrete situations to use normalized weighted gamma distribution \( D(\alpha_1, ..., \alpha_l; \beta) \).

3.2 Ferguson and Lo random distributions

Let \( \alpha \) be a finite measure on \( V \). A random distribution \( X : \Omega \to P(V) \) is a Dirichlet process \( D(\alpha) \) if for every \( k = 1, 2, 3, ... \) and every measurable partition \( B_1, ..., B_k \) of \( V \), the joint distribution of the random vector \((X(B_1), ..., X(B_k))\) is a Dirichlet distribution with parameters \((\alpha(B_1), ..., \alpha(B_k))\), [11].

Lo [16] has similarly defined a normalized weighted gamma distribution.

All such RDs are discrete, that is \( X(\omega) \) is a discrete distribution for a.a. \( \omega \). In the following section we study a continuous RD.

4 A continuous random distribution derived from Haar wavelets

Kraft, [15], defined a random distribution \( X \) such that the probability measure \( X(\omega) \) has a density w.r.t. the Lebesgue measure on \( V = [0, 1] \). In this section, we show that:

- we don’t need some conditions on the independance and the expectation of the parameters.
- Kraft random distribution can be considered as a randomized Haar wavelet serie.
- the construction can be extended on \( V = [0, 1]^2 \) for image purposes.

This enlightens Kraft method and can be of interest for further constructions derived from other wavelets. Some of the proofs below are given very shortly in Kraft’s paper. For completeness, we precise here all the arguments.
4.1 Kraft random distribution on \( V = [0, 1] \)

4.1.1 Construction

Kraft’s construction hinges on a set
\[
Z = \{Z_{\frac{k}{2^r}} \text{ with } r = 1, 2, \ldots \text{ and } k = 1, 3, \ldots 2^r - 1\}
\]
of random variables defined on \((\Omega, \mathcal{F}, P)\) such that
\[
0 < Z_{\frac{k}{2^r}} < 1.
\]

Note that we do not assume the independance of \(Z_i\) as in Kraft’s paper.

Let \((F_l)_{l \geq 1}\) be a sequence of random cumulative distribution functions (cdf) on \([0, 1]\) defined by induction as follows:
\[
F_1(0) = 0, \quad F_1\left(\frac{1}{2}\right) = Z_{\frac{1}{2}}, \quad F_1(1) = 1
\]

\(F_1\) is affine on \([0, \frac{1}{2})\) and \([\frac{1}{2}, 1]\).

For any \(l = 1, 2, \cdots\) and any odd \(k\) from 1 to \(2^l - 1\),
\[
F_l\left(\frac{k}{2^l}\right) = F_{l-1}\left(\frac{k-1}{2^l}\right)(1 - Z_{\frac{k}{2^l}}) + F_{l-1}\left(\frac{k+1}{2^l}\right)Z_{\frac{k}{2^l}},
\]
for any even \(k\) from 0 to \(2^l\),
\[
F_l\left(\frac{k}{2^l}\right) = F_{l-1}\left(\frac{k}{2^l}\right),
\]
and for any \(k = 0, ... 2^l - 1\) \(F_l\) is affine on \([\frac{k}{2^l}, \frac{k+1}{2^l}]\).

The random distribution \(X\) is then defined by its random cdf \(F\) as follows:
\[
F\left(\frac{k}{2^l}\right) = F_l\left(\frac{k}{2^l}\right),
\]
and for any \(x \in [0, 1]\)
\[
F(x) = \text{sup}( F(r) : r < x, r \text{ dyadic rational}).
\]

Note that \(F\) is strictly increasing. The following can be proved as in Gerald [9]:

**Proposition 4.1** If the \(Z_{\frac{k}{2^r}}\) for any \(l\) and any \(0 \leq k \leq 2^l\) are identically distributed with arbitrary density function \(f\) such that the support of \(f\) is \([0, 1]\), then the random cdf \(F\) is continuous with probability one.

4.1.2 Polya tree structure

The random function \(F_l\) has, except at \(\frac{k}{2^l}\), a random derivative \(g_l\) which is constant on the dyadic intervals \(I_{l,k} = (\frac{k}{2^l}, \frac{k+1}{2^l})\) for any \(l\) and \(k = 0, ... 2^l - 1\).

If \(\lambda\) denotes the Lebesgue measure on \([0, 1]\) and \(\mathcal{F}_l\) the finite \(\sigma\)-algebra generated by the dyadic intervals \((I_{l,k})_{0 \leq k < 2^l}\), then observe that the mapping
\[ g_l: (\Omega \times [0, 1], \mathcal{F} \otimes \mathcal{F}_l, P \otimes \lambda) \to \mathbb{R} \]

\[ (\omega, x) \quad \to \quad g_l(\omega, x) \]

is measurable.

Moreover, for any \( \omega \in \Omega \), \( g_l(\omega, .) \) is a density:

\[ \int_0^1 g_l(\omega, x)dx = 1. \]

and if \( x \in [0, 1] \) is written as

\[ x = \sum_{r=1}^{\infty} \frac{\epsilon_r(x)}{2^r} \quad \text{with} \quad \epsilon_r(x) \in \{0, 1\} \]

and \( k_r(x) \) is the integer defined by

\[ \frac{k_r(x)}{2^r} \leq x < \frac{k_r(x) + 1}{2^r}, \]

then we have ([15])

\[ g_l(\omega, x) = 2^l \prod_{r=1}^{l} Z_{k_r(x)+1}^{(1-\epsilon_r(x))}(\omega)(1 - Z_{k_r(x)}^{(1-\epsilon_r(x))}(\omega))^{\epsilon_r(x)}. \]  

This expression can be obtained through a Polya tree structure, [18], described below for \( l = 3 \) :

\[ \text{Figure 1: Upper levels of a Polya tree structure described for } l = 3. \text{ For example if } x \in [3/8, 1/2], g_3(\omega, x) = Z_{1/8}^1(\omega)(1 - Z_{1/4}^1(\omega))(1 - Z_{3/8}^1(\omega)) \text{ for any } \omega \in \Omega \]
4.1.3 Convergence of the sequence $g_l$

Kraft [15] has proved that if there exists a constant $K < \infty$ such that for every $l$,

$$
\int_0^1 \int_\Omega g_l^2(\omega, x) dx dP < K \tag{2}
$$

then with probability one

1. $\lim_{l \to +\infty} g_l(\omega, x) = g(\omega, x)$ for almost all $x \in [0, 1]$ and $\omega \in \Omega$,

2. $\lim_{l \to +\infty} F_l(\omega, x) = F(\omega, x)$ and $F(\omega, x) = \int_0^x g(\omega, t) dt, \ 0 \leq x \leq 1$,

3. $g \in L^2(\Omega \times [0, 1]).$

In the next sections, we will use the following notation

$$
g = Kraft(Z).
$$

The proof of the previous points is based on the following result

**Lemma 4.2** $(g_l)_l$ is a martingale w.r.t. the filtration $(\mathcal{F} \otimes \mathcal{F}_l)_l$ and the measure $P \otimes \lambda$.

**Proof of the Lemma**

We have to show that :

(a) $g_l$ is $\mathcal{F} \otimes \mathcal{F}_l$ measurable,

(b) $g_l = E(g_{l+1}|\mathcal{F} \otimes \mathcal{F}_l)$ i.e. $\forall A \in \mathcal{F} \otimes \mathcal{F}_l, \int_A g_l = \int_A g_{l+1}$.

Let $B$ be a borelian set on $\mathbb{R}$ and let

$$
g_l^{-1}(B) = \{(\omega, x) \in \Omega \times [0, 1] \text{ such that } g_l(\omega, x) \in B\}
$$

so that

$$
g_l^{-1}(B) = \bigcup_{0 \leq k < 2^l} \{\omega \in \Omega \times I_{l,k}, \ g_l(\omega, x) \in B\} \times \{x\}.
$$

As $g_l(\omega, .)$ is constant on $I_{l,k}$, we have

$$
g_l^{-1}(B) = \bigcup_{0 \leq k < 2^l} \{\omega, \ g_l(\omega, \frac{k}{2^l}) \in B\} \times I_{l,k}.
$$

Due to (1), we see that for any $x \in [0, 1], \ g_l(., x)$ is measurable w.r.t. $\mathcal{F}$. Therefore $\{\omega, \ g_l(\omega, \frac{k}{2^l}) \in B\}$ belongs to $\mathcal{F}$ and $g_l^{-1}(B)$ belongs to $\mathcal{F} \otimes \mathcal{F}_l$. This proves point a).

To prove point b), observe that for any $\omega \in \Omega$

$$
\int_{I_{l,k}} g_l(\omega, x) dx = F_l(\frac{k + 1}{2^l}) - F_l(\frac{k}{2^l})
$$

6
\[ \int_{I_{l,k}} g_{l+1}(w, x)dx = F_{l+1}(2k + 1) \frac{1}{2^{l+1}} - F_{l+1}(2k + 2) \frac{1}{2^{l+1}} + F_{l+1}(2k + 1) \frac{1}{2^{l+1}} - F_{l+1}(2k + 1) \frac{1}{2^{l+1}}. \]

So
\[ \int_{I_{l,k}} g_{l+1}(w, x)dx = F_{l+1}(k + 1) \frac{1}{2^{l}} - F_{l+1}(k) \frac{1}{2^{l}}. \]

Therefore we have
\[ \int_{I_{l,k}} g_{l}(w, x)dx = \int_{I_{l,k}} g_{l+1}(w, x)dx \]
so that Fubuni’s theorem implies point b) for any \( A \in \mathcal{F} \otimes \mathcal{F}_l \) when \( A = B \times I_{l,k} \). This yields point b) for general \( A \in \mathcal{F} \otimes \mathcal{F}_l \).

\[ \square \]

Now we prove the three points:

1. We apply the martingale convergence theorem, Doob [8]. In order to get the a.s. convergence, we need
   \[ \sup_l (E(|g_l|)) < \infty. \]
   We get
   \[ E(|g_l|) = \int_0^1 \int_\Omega g_l(\omega, x)dx dP, \]
   Since for any \( \omega \), \( g_l(\omega, x) \) is a density function:
   \[ \int_0^1 g_l(\omega, x)dx = 1 \]
   therefore
   \[ E(|g_l|) = \int_\Omega 1dP = P(\Omega) = 1. \]

2. We prove that \( g_l \) is uniformly integrable:
   \[ \int g_l > M g_l dPd\sigma = \int_0^1 (\int_\Omega g_l > M d\sigma)d\sigma \]
   \[ \leq \int_0^1 (\int_\Omega g_l^2 d\sigma) \frac{1}{2} (\int_\Omega g_l > M)_l d\sigma \]
   \[ \leq \int_0^1 (\int_\Omega g_l^2 d\sigma) \frac{1}{2} (P(\omega, g_l(\omega, x) > M)) d\sigma \]
   \[ \leq \int_0^1 (\int_\Omega g_l^2 d\sigma) \frac{1}{2} \frac{E(g_l)}{M} d\sigma \]
   \[ \leq \int_0^1 (\int_\Omega g_l^2 d\sigma) \frac{1}{2} \frac{1}{M} d\sigma \]
   \[ \leq \int_0^1 (\int_\Omega g_l^2 d\sigma) \frac{1}{2} \frac{1}{M} d\sigma \]
   We apply the martingale convergence theorem, Doob [8], we deduce the convergence in \( L^1(\Omega \times [0, 1]) \).

3. With Fatou’s lemma and Fubini’s theorem, we have:
   \[ \int_\Omega \int_0^1 g^2 dPdx \leq \lim \int_\Omega \int_0^1 g_l^2 dPdx \leq K. \]
The following proposition shows how the control of the variance of the $Z_{k/2^l}$ implies condition (2).

**Proposition 4.3** If $Z$ is a set of completely independent real random variables such that for any $l$ and any $0 \leq k < 2^l$ :

$$E(Z_{k/2^l}) = \frac{1}{2}$$

and

$$\sup_{0<k<2^l, k \text{ odd}} \text{var}(Z_{k/2^l}) \leq b_l \text{ and } \sum b_l < \infty$$

then condition (2) is satisfied.

In particular, this is the case if there exists a constant $C$ such that for any $l$ and any $0 \leq k < 2^l$ :

$$E(Z_{k/2^l}) = \frac{1}{2},$$

$$2^l \text{var}(Z_{k/2^l}) \leq C.$$

**Proof of proposition 4.3**

We have

$$E_P(g_l^2(\omega, x)) = \prod_{r=1}^l 4E((Z_{k_{r}(x)+1}^2(1-\epsilon_r(x))((1 - Z_{k_r(x)/2^l}(\omega))^2)^\epsilon_r(x))$$

because the random variables $Z_i$ are independent.

From the hypothesis, we get

$$E(Z_{k_{r}/2^l}) \leq b_r + \frac{1}{4},$$

and

$$E((1 - Z_{k_{r}/2^l})^2) \leq b_r + \frac{1}{4}.$$  

Let $x \in [0, 1]$. Since $\epsilon_r(x)$ is either 1 or 0, we have

$$E_P(g_l^2(\omega, x)) \leq \prod_{r=1}^l (4b_r + 1).$$

Since $\prod_{r=1}^l (4b_r + 1)$ converges if $\sum b_r < \infty$, let

$$K = \prod_{r=1}^\infty (4b_r + 1),$$

so that

$$E_P(g_l^2(\omega, x) \leq K$$

8
Thus condition (2) holds proving the existence of \( g = Kraft(Z) \) with

\[
g(\omega, x) = \prod_{r=1}^{\infty} 2Z_{k_r/2}^{(1-\epsilon_r(x))}(\omega)(1 - Z_{k_r/2}^{\epsilon_r(x)}). \tag{3}
\]

\[\square\]

Remark: It can be seen that for any odd \( k \)

\[
Z_{k/2} = \frac{F(k/2) - F(k-1/2)}{F(k+1/2) - F(k/2)}
\]
so that the coefficient \( Z_{k/2} \) can be interpreted as a conditionnal probability w.r.t. \( \lambda : \)

\[
Z_{k/2}(\omega) = \lambda(X(\omega, \cdot) \in I_{l, k-1} | X(\omega, \cdot) \in I_{l-1, (k-1)/2})
\]
for any \( \omega \in \Omega \).

### 4.1.4 A Posteriori Distribution

The random distribution \( F \) is uniquely determined by the finite-dimensional distributions of the vectors \( F(I_{l, k}) \) with \( l > 0, k = 0, \ldots, 2^l - 1 \). Thus if we suppose that \( F \) is uncertain, a prior distribution for \( F \) can be defined through a prior beta distribution for the set of Kraft coefficients:

\[
Z_{k/2} \sim \beta_{a_{kl}, b_{kl}}, \; l = 1, 2, \ldots; \; k = 1, 3, \ldots, 2^l - 1
\]

Recall that the density of a beta distribution is defined as:

\[
\beta_{a, b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}1_{[0,1]}(x)
\]

where \( 1_{[0,1]} \) is the indicator function of the set \([0,1]\), with mean and variance equal to \( \frac{a}{a+b} \) and \( \frac{ab}{(a+b+1)(a+b)^2} \) respectively.

If we observe the data \( x_i \in [0,1] \), it is well known that the posterior distribution of \( Z_{k/2} \) is:

- \( Z_{k/2}|x_i \sim \beta(a_{kl} + 1, b_{kl}) \) if \( x_i \in \left[\frac{k-1}{2^l}, \frac{k}{2^l}\right] \),
- \( Z_{k/2}|x_i \sim \beta(a_{kl}, b_{kl} + 1) \) if \( x_i \in \left[\frac{k}{2^l}, \frac{k+1}{2^l}\right] \),

Since the marginal distributions of a conditionnal distribution are equal to the conditionnal distributions of the marginals, we get the commutativity of the following diagram for a given \( l \)

\[
g_l = Kraft(Z) \quad \rightarrow \quad Z_{k/2} \sim \beta(a_{kl}, b_{kl}),
\]

\[
g_l|x_i \quad \rightarrow \quad Z_{k/2} \sim \beta(a'_{kl}, b'_{kl})
\]

However given a prior \( g = Kraft(Z) \), it is not sure because of condition (2) that the posterior \( g|x_i \) is still a Kraft process.
4.1.5 Simulation

Kraft random distribution can be used as a generator of random distributions. If \( a_{kl} = b_{kl} = 2^{l-3} \) for any \( k \) and any \( l \), then proposition 4.3 shows that \( F \) is absolutely continuous and that

\[
g = Kraft(Z) .
\]

In figure 2, we display some simulations of the function \( g_l \) with \( l = 7 \) in two cases: in the first one, the conditions of proposition 4.3 are not satisfied with \( a_{kl} = b_{kl} = 0.2 \) for any \( k \) and any \( l \), so that the cdf \( F \) need not to be continuous. In the second case we set \( a_{kl} = b_{kl} = 2^{l-3} \) for any \( k \) and \( l \), so that the cdf \( F \) is continuous according to proposition 4.3.

![Figure 2: In the left column, we represent 3 distributions with beta distribution \( a_{kl} = b_{kl} = 0.2 \) and in the right column we represent 3 distributions with beta distribution \( a_{kl} = b_{kl} = 2^{l-3} \).](image)

4.1.6 Haar wavelet

In 1909, Alfred Haar gave a construction of an orthonormal basis of \( L^2([0,1]) \). It is composed of the function \( \phi \) and of the \( \psi_{j,k} \) defined by

\[
(\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k))_{j \in \mathbb{N}, k=0,...,2^j-1} .
\]

The function \( \psi \) is defined in the following way:

\[
\psi(x) = \begin{cases} 
1 & \text{if } x \in \left[0, \frac{1}{2}\right) \\
-1 & \text{if } x \in \left[\frac{1}{2}, 1\right) \\
0 & \text{otherwise}
\end{cases}
\]
while the scale function $\phi$ is defined as:

$$\phi(x) = 1 \quad \forall x \in [0, 1]$$

For any $g \in L^2([0, 1])$, we have:

$$g = a\phi + \sum_{j=0}^{+\infty} \sum_{k=0}^{2^j-1} W_{j,k}(g) \psi_{j,k}.$$  

where the limit is taken in $L^2([0, 1])$. Since $\int_0^1 \psi_{j,k}(x) dx = 0$, we see that $a = \int_0^1 g(x) dx$.

The wavelet coefficient $W_{j,k}(g)$ is computed with the scalar product of $g$ with $\psi_{j,k}$:

$$W_{j,k}(g) = \int_0^1 g(x) \psi_{j,k}(x) dx.$$  

For a given $J$, we set:

$$g_{J}^{\text{haar}} = a\phi + \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} W_{j,k}(g) \psi_{j,k}.$$  

The following result establishes the link between Kraft construction and Haar wavelet decomposition.

**Proposition 4.4** Let $g = Kraft(Z)$ be defined as in (3) and $g_l^{\text{kraft}}$ be defined as in (1). Let $W_{j,k}$ be the Haar wavelet coefficients of $g \in L^2([0, 1])$. For any $x \in [0, 1]$ let

$$g_{J}^{\text{haar}}(\omega, x) = 1 + \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} W_{j,k}(g, \omega) \psi_{j,k}(x).$$  

Then we have

$$g_l^{\text{kraft}} = g_{l-1}^{\text{haar}} \text{ for any } l \geq 1.$$  

**Proof of Proposition 4.4**

Step 1 : We have

$$\int_0^1 g(t) 1_{I_{J+1,k}}(t) dt = \int_0^1 g_{J}^{\text{haar}}(t) 1_{I_{J+1,k}}(t) dt.$$  

As $g_{J}^{\text{haar}}$ is constant on the dyadic interval $I_{J+1,k}$, it is easy to check that for any $x \in I_{J+1,k}$

$$g_{J}^{\text{haar}}(x) = 2^{J+1} \int_{I_{J+1,k}} g(t) dt.$$  

(4)

Step 2 : On the other hand, the function $g_l^{\text{kraft}}$ is constant on the dyadic interval $I_{l,k}$ for any $0 \leq k \leq 2^l - 1$. As $g_l^{\text{kraft}}$ is a martingale (lemma 4.2), we have for any $x \in I_{l,k}$:

$$g_l^{\text{kraft}}(x) = 2^l \int_{I_{l,k}} g(t) dt.$$  

11
Step 3: So we deduce:

\[ g_l^{kraft} = g_{l-1}^{haar} \text{ for any } l \geq 1. \]

\[ \square \]

**Corollary 4.5** Let \( g = Kraft(Z) \), if condition (2) is satisfied, then for almost all \( \omega \in \Omega \), \( g_l^{kraft}(\omega, \cdot) \) converge to \( g(\omega, \cdot) \) in \( L^2([0,1]) \).

**Proof of Corollary 4.5**

As \( g_J^{haar} \) converges to \( g(\omega, \cdot) \) in \( L^2([0,1]) \), the result is a consequence of proposition 4.4.

\[ \square \]

### 4.1.7 Relation between Kraft and Haar coefficients

We can express the Kraft coefficients as function of wavelet coefficients:

Indeed, for any \( x \in \left( \frac{k-1}{2}, \frac{k}{2} \right) \) with \( k \) odd:

\[
Z_{\frac{k}{2^l}} = \frac{1}{2} \frac{g_l^{kraft}(x)}{g_{l-1}^{kraft}(x)},
\]

we have

\[
Z_{\frac{k}{2^l}} = \frac{1}{2} \frac{g_l^{haar}(x)}{g_{l-2}^{haar}(x)},
\]

and therefore

\[
Z_{\frac{k}{2^l}} = \frac{1 + \sum_{j=0}^{l-1} 2^{j/2} W_{j,k}(x)}{2(1 + \sum_{j=0}^{l-2} 2^{j/2} W_{j,k}(x))},
\]

(5)

Conversely, we have an expression of the wavelet coefficients as function of Kraft coefficients: since

\[
W_{l,k} = \int_0^1 g(x) \psi_{l,k}(x) dx,
\]

we have

\[
W_{l,k} = 2^{l/2} \left( \int_{I_{l+1,2k}} g(x) dx - \int_{I_{l+1,2k+1}} g(x) dx \right),
\]

therefore for any \( x_1 \in I_{l+1,2k} \) and \( x_2 \in I_{l+1,2k+1} \), we have

\[
W_{l,k} = 2^{-l/2} (g_{l+1}^{kraft}(x_1) - g_{l+1}^{kraft}(x_2)).
\]

We deduce that for any \( x = x_1 \) or \( x_2 \):

\[
W_{l,k} = 2^{-l/2} g_l^{kraft}(x) (2Z_{\frac{k}{2^{l+1}}} - 1),
\]

so that

\[
W_{l,k} = 2^{l/2} \prod_{r=1}^l Z_{\frac{k_r(x)}{2^{r+1}}}^{(1-\epsilon_r(x))} (1 - Z_{\frac{k_r(x)}{2^r}})^{\epsilon_r(x)} (2Z_{\frac{k_r(x)}{2^{r+1}}} - 1).
\]

(6)
4.1.8 Random distributions derived from Haar wavelets

Given a family \( W = (W_{l,k}) \) of random Haar wavelet coefficients, we can define a family \( Z = (Z_{l,k}) \) with (5), but we are not sure that the coefficients \( Z_{l,k} \) belong to \([0, 1]\). So we give another approach to get random distributions from Haar wavelets. Apply the logarithm to both terms in equation (1) to get:

\[
Ln(g_l(\omega, x)) = lLn(2) + \sum_{r=1}^{l} (1 - \epsilon_r(x)) Ln(Z_{r^2, (2r+1)/2}) + \epsilon_r(x)Ln(1 - Z_{r^2, (2r+1)/2})
\]

then using indicator functions, we have

\[
Ln(g_l(\omega, x)) = lLn(2) + \sum_{r=1}^{l} \sum_{k=0}^{2^{r-1} - 1} \frac{1}{2^{(1+r)/2}} Ln\left(\frac{Z_{r^2, (2r+1)/2}}{1 - Z_{r^2, (2r+1)/2}}\right) \psi_{r-1,k}(x).
\]

Let define

\[
A_l = \int_0^1 Ln(g_l(\omega, x))dx = lLn(2) + \sum_{r=1}^{l} \sum_{k=0}^{2^{r-1} - 1} \frac{1}{2^{r}} Ln\left(\frac{Z_{2^{r+1}, (2^{r+1}+1)/2}}{1 - Z_{2^{r+1}, (2^{r+1}+1)/2}}\right),
\]

so that

\[
Ln(g_l(\omega, x)) = A_l + \sum_{r=1}^{l} \sum_{k=0}^{2^{r-1} - 1} \frac{1}{2^{(1+r)/2}} Ln\left(\frac{Z_{2^{r+1}, (2^{r+1}+1)/2}}{1 - Z_{2^{r+1}, (2^{r+1}+1)/2}}\right) \psi_{r-1,k}(x).
\]

Then the wavelet coefficients of \( Ln(g_l(\omega, x)) \) are given by

\[
W_{r-1,k} = \frac{1}{2^{(1+r)/2}} Ln\left(\frac{Z_{2^{r+1}, (2^{r+1}+1)/2}}{1 - Z_{2^{r+1}, (2^{r+1}+1)/2}}\right),
\]

for \( r \geq 1 \).

Conversely we see that

\[
Z_{2^{r+1}, (2^{r+1}+1)/2} = \frac{e^{2(1+r)/2W_{r-1,k}}}{1 + e^{2(1+r)/2W_{r-1,k}}} \in [0, 1].
\]

So we have proved the following result

**Proposition 4.6** Let \( W = (W_{r,k}) \) be a set of real random variables defined on \((\Omega, F, P)\).

For any \( l \), let define:

\[
f_l(\omega, x) = A_l + \sum_{r=1}^{l} \sum_{k=0}^{2^{r-1} - 1} W_{r-1,k} \psi_{r-1,k}(x),
\]

with

\[
A_l = lLn(2) + \sum_{r=1}^{l} \sum_{k=0}^{2^{r-1} - 1} \frac{1}{2^{r}} Ln\left(\frac{e^{2(1+r)/2W_{r-1,k}}}{1 + e^{2(1+r)/2W_{r-1,k}}}\right).
\]

Then the function \( g_l = e^{f_l} \) is a Kraft random distribution with respect to the family \( Z = (Z_{2^{r+1}, (2^{r+1}+1)/2}) \) with

\[
Z_{\frac{2^{r+1}}{2}} = \frac{e^{2(1+r)/2W_{r-1,k}}}{1 + e^{2(1+r)/2W_{r-1,k}}}.
\]
4.2 Case $V = [0, 1] \times [0, 1]$

In this subsection, the connection between Kraft construction and Haar wavelet is extended to dimension 2.

In [18], Kraft process are defined on $[0, 1] \times [0, 1]$.

Their construction hinges on a set

$$Y = (Y_{\epsilon_1...\epsilon_i...})$$

of real random variables on $[0, 1] \times [0, 1]$ with $\epsilon_i \in \{0, 1, 2, 3\}$ such that

$$Y_{\epsilon_1...\epsilon_i0} + Y_{\epsilon_1...\epsilon_i1} + Y_{\epsilon_1...\epsilon_i2} + Y_{\epsilon_1...\epsilon_i3} = 1.$$

Consider a dyadic partitioning of $[0, 1] \times [0, 1] :$

$$I_{l,k} = \left[ \frac{k_1}{2^l}, \frac{k_1 + 1}{2^l} \right] \times \left[ \frac{k_2}{2^l}, \frac{k_2 + 1}{2^l} \right] \text{ with } k = (k_1, k_2) \in \{0, ..., 2^l - 1\}^2.$$

For any point $x = (x_1, x_2) \in [0, 1] \times [0, 1]$, consider a sequence $(\epsilon_n)_{n \geq 0}$ in $\{0, 1, 2, 3\}$ such that $\epsilon_n = \epsilon^i_n + 2\epsilon^2_n$ where $\epsilon^i_n = \epsilon_n(x_i)$.

Figure 3 gives an illustration for $l = 1$.

![Figure 3: Definition of $\epsilon$.](image)

Let :

$$g^{kraft}_l(x) = 2^l \prod_{i=1}^{l} Y_{\epsilon_1...\epsilon_i}.$$ 

The function $g^{kraft}_l$ is constant on $I_{l,k}$ and the coefficient $Y$ can be interpreted as a conditionnal probability.

It is easy to check that $(g^{kraft}_l, \mathcal{F} \otimes \mathcal{F}_l)_l$ is a martingale w.r.t. $P \otimes \lambda$. Kraft’s construction can be generalized to define

$$g = kraft2(Y).$$

An example of such process can be constructed by letting the random vector $Y$ following a Dirichlet distribution $D(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

Using bidimensionnal Haar wavelet, it is straightforward to generalize proposition 4.4 in dimension 2.

5 Mixtures of random distributions

In practice and specially in image analysis, mixture of random distributions are more appropriate to modelize concrete situations.
5.1 The finite mixture problem

Let \((x_1, \ldots, x_N)\) be \(N\) observations of a sample of size \(N\) from a random distribution \(X\). The problem consists in estimating the distribution \(\mathcal{P}_X\) of \(X\) when \(\mathcal{P}_X\) is supposed to be a convex combination of distributions \(P_k\) belonging to a specific family:

\[ \mathcal{P}_X = \sum_{k=1}^{K} p_k P_k, \]

with \(p_k > 0\) and \(\sum_{k=1}^{K} p_k = 1\).

5.2 Mixture of Dirichlet Distributions

5.2.1 Estimation in finite dimension

We consider here that \(X\) is a random variable:

\[ X : (\Omega, \mathcal{F}, \mathcal{P}) \rightarrow E_l \]

\[ \omega \mapsto X(\omega) \]

where

\[ E_l = \{ y = (y_1, \ldots, y_l), y_j \geq 0, \sum_{j=1}^{l} y_j = 1 \}. \]

and we assume that \(\mathcal{P}_X\) is a mixture of \(K\) standard Dirichlet distributions, that is

\[ \mathcal{P}_X = \sum_{k=1}^{K} p_k D(\alpha_1^k, \ldots, \alpha_l^k) \]

with \(p_k > 0\) and \(\sum_{k=1}^{K} p_k = 1\).

Given \(N\) observations of \(X\), that is \(N\) \(l\)-uples \(y_i = (y_{i,1}, \ldots, y_{i,l})\), \(y_i\) belonging to \(E_l\), the mixture problem consists in estimating the parameter

\[ \Theta = (p_k, (\alpha_{1}^k, \ldots, \alpha_{l}^k))_{1 \leq k \leq K}. \]

In the following steps, we modify the estimation step of SAEM algorithm [5], a stochastic variant of the popular E.M. algorithm [6].

5.2.2 Description of the algorithm

The inputs are \(N\) vectors \(y_i = (y_{i,1}, \ldots, y_{i,l}) \in \mathbb{R}^l, i = 1, 2, \ldots, N\) and the number of components is a given integer \(K\).

It is an iterative algorithm:

- Initialization : start with \(\Theta_0\).
- Iteration : let \(\Theta_n\) be the parameters obtained at step \(n\),

\[ \Theta_n = (p_{k,n}, (\alpha_{1,n}^k, \ldots, \alpha_{l,n}^k))_{1 \leq k \leq K}. \]
1. Compute the coefficients $t_{ik}^n$:

$$t_{ik}^n = \frac{p_{k,n} \mathcal{D}(\alpha_{1,n}^k, \ldots, \alpha_{l,n}^k)(y_i)}{\sum_{h=1}^K p_{h,n} \mathcal{D}(\alpha_{1,n}^h, \ldots, \alpha_{l,n}^h)(y_i)}.$$

2. Stochastic step: generate multinomial numbers $e_i^n$

$$e_i = (e_{ik}^n)_{k=1, \ldots, K}$$

of one draw of $K$ categories with probabilities $(t_{ik}^n)_{k=1, \ldots, K}$ so that all the $e_{ik}^n$ are 0 except one of them equal to 1. We then get a partition $C = (C_k)_{k=1, \ldots, K}$ of the sample by letting

$$C_k = \{y_i \in \mathbb{R}^l : e_{ik}^n = 1\}.$$

If the cardinality of $C_k$ for some $k$ is too small, $< \delta$, then repeat the algorithm with a new initialisation.

3. Estimation step: estimate the parameters $p_{k,n+1}$ and $\alpha_{j,n+1}$ in each class given by the stochastic step. Then we use the following result: if $Y = (Y_1, \ldots, Y_l) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_l)$, we have:

$$E(Y_j) = \frac{\alpha_j}{\alpha} V(Y_j) = \frac{\alpha_j(1 - \frac{\alpha_j}{\alpha})}{\alpha + 1}$$

where $\alpha = \alpha_1 + \ldots + \alpha_l$. Hence we get:

$$\alpha = \frac{E(Y_j)(1 - E(Y_j))}{V(Y_j)} - 1$$

$$\alpha_j = \alpha E(Y_j)$$

So we deduce the estimation step for the class $C_k$

$$p_{k,n+1} = \frac{\text{Card}(C_k)}{N} = \frac{\sum_{i=1}^N e_{ik}^n}{N}$$

$$\alpha_{k,n+1} = \frac{\hat{m}_{j,l}(1 - \hat{m}_{j,l})}{\hat{\sigma}_{j,l}^2} - 1$$

$$\alpha_{j,n+1}^k = \alpha_{k,n+1} \hat{m}_{j,l}$$

where $\hat{m}_{j,l}$ and $\hat{\sigma}_{j,l}^2$ are respectively estimators of $E(Y_{j,l})$ and $V(Y_{j,l})$.

A similar algorithm can be used to estimate a finite mixtures of Kraft random distributions defined as

$$X(\omega) = \sum_{k=S}^{k} 1_{U=U_k}(\omega) \text{Kraft}(Z^s)(\omega),$$

with $p_k > 0$ and $\sum_{k=1}^K p_k = 1$ and where $U$ is a discrete random variable such that $P(U = U_k) = p_k$ and $U$ is independent from the Kraft distributions.

Using the mutual singularity of Dirichlet (resp.Kraft) RDs, it can be proved that such algorithms yield consistent estimators of the parameters of the mixture ($[10]$).
6 Applications in image analysis

6.1 A new hierarchical model of image

Our model is based on the following hierarchical scheme: let given positive \( \alpha_{1,l}^{k,l}, \ldots, \alpha_{4,l}^{k,l} \) with \( l \geq 1 \) and \( 0 \leq k \leq 4^l \). For \( i = 1, 2, \ldots \), let \( \epsilon_i \in \{0, 1, 2, 3\} \) and let

\[
(Y_{\epsilon_1 \ldots \epsilon_i0}, Y_{\epsilon_1 \ldots \epsilon_i1}, Y_{\epsilon_1 \ldots \epsilon_i2}, Y_{\epsilon_1 \ldots \epsilon_i3}) \sim \mathcal{D}(\alpha_1^{k,l}, \ldots, \alpha_4^{k,l})
\]

Let \( Y = (Y_{\epsilon_1 \ldots \epsilon_i}) \) and \( g = \text{kraft2}(Y) \) the bidimensional random distribution on \([0, 1] \times [0, 1]\) as defined in subsection 4.2.

Let \( P|g \sim g \) be a probability measure on \([0, 1] \times [0, 1]\) generated by \( g \).

Let \( X_1, \ldots, X_n|P \sim \text{iid} P \) be \( n \) pixels generated by \( P \).

In figure 4, we display two densities \( P \) generated by a bidimensional Kraft random distribution. Each of these densities generates two black and white images of \( n \) points. In the next figure, we apply the hierarchical model with a mixture of two Kraft RDs.

We think that this kind of model can be of interest for texture analysis and classification of images.

6.2 Segmentation based on local distributions

Usually, segmentation gather pixels with similar grey level. We propose here a segmentation based on similar local distributions. By local distribution, we mean distribution of grey level or curvature within a neighbourhood of each pixel.

In these experiments, we consider a grey level image. At each pixel, the grey level is a number between 0 (black) and 255 (white).
Our segmentation algorithm is based on both estimation of Dirichlet mixtures and the Polya urn with temporal and spatial contagion scheme:

- initialization

1. histogram sampling: select $N$ pixels randomly or not. For each of them, compute within a neighborhood a frequency grey level histogram with $l$ bins. This yields $N$ probability vectors.
2. estimation of the Dirichlet mixture: apply algorithm 4.2.2 to these vectors in order to estimate a mixture of $K$ Dirichlet distributions.
3. posterior probabilities: for all pixel $i$, compute the grey level histogram $y_i = (y_{ij})_{1 \leq j \leq l}$, compute the probability $t_{ik}$ that pixel $i$ belongs to class $k$

$$t_{ik} = \frac{p_k D(\alpha^k_1, \ldots, \alpha^k_l)(y_i)}{\sum_{h=1}^{K} p_h D(\alpha^h_1, \ldots, \alpha^h_l)(y_i)}$$

4. Polya urn: each pixel $i$ is associated to an urn with a mixture of balls of $K$ different colors: one color for each class label. The proportion of balls of color $k$ is $t_{ik}$ for all $1 \leq k \leq K$.

- iterative step

At time $n$, the urn composition of each pixel $i$ is updated by combining the balls of the neighboring urns with the balls of the pixel $i$ at time $n-1$. So we get a super urn from which a ball is sampled. Then $\Delta$ balls of that color are added to the urn of pixel $i$.

We made some experiments, the result of which is shown in figure 4.

7 Conclusion

We have introduced random distributions in image modelling and analysis. We have mainly used Dirichlet RDs and Kraft RDs, these last being considered as randomized Haar wavelets. We proposed a hierarchical model and a segmentation algorithm based on such RDs. We think that this approach can be further developed in both theoretical and practical aspects for classifying and mining huge image databases.
Figure 6: Above: the initial image, fingerprint, in the middle: the initial image in the segmentation algorithm, Below: the segmentation.

References


