A GENERALISATION OF THE MIXTURE DECOMPOSITION PROBLEM IN
THE SYMBOLIC DATA ANALYSIS FRAMEWORK

E. Diday 23/4/2001

Summary
In Symbolic Data Analysis, more complex units can be considered like "concepts" (as towns, insurance companies, species of animals). A concept can be characterized by an "extent" defined by a class of standard units called "individuals" (as a sample of inhabitant of a town, a sample of insurance companies, a sample of animals of a given species). These classes can be described by a distribution associated to each variable, summarizing in that way huge sets of data. Therefore, here we are interested by the case where each unit representing a "concept" is described by a vector of p distributions associated to p variables. Our aim is to find simultaneously a "good" partition of these units and a model using "copulas" associated to each class of this partition. Different copulas models are recalled where the case of Markov process and Brownian motion are considered. The mixture decomposition problem is settled in this general case. It extends the standard mixture decomposition problem to the case where each unit is described by a vector of distributions instead as usual, by a vector of unique (categorical or numerical) values. Several generalization of standard algorithms are suggested. One of them is illustrated by a simple example. All these results are first considered in the case of a unique variable and then extended to the case of a vector of p variables by using a top-down binary tree approach. Finally, the case of infinite joint and copulas is considered.

Key-words: Mixture decomposition, Symbolic Data Analysis, Data Mining, Clustering, Partitioning

1. Introduction
In a symbolic data table, a cell can contain, a distribution (Schweitzer (1984) says that "distributions are the number of the future"!), or intervals, or several values linked by a taxonomy and logical rules, etc.. The need to extend standard data analysis methods (exploratory, clustering, factorial analysis, discrimination,...) to symbolic data table is increasing in order to get more accurate information and summarise extensive data sets by the description of the underlying concepts contained in Data Bases (as towns or socio-economic groups) considered as new kinds of units.

The idea of operating with distribution functions as data is already explicit in Mengers (1949, 1970) writings and applied in clustering by Janowitz and Schweitzer (1989). We are here interested to extend the mixture decomposition problem (as defined for instance in Dempster and al (1977)) to the case where the units are described by distributions.

Here a "variable" is considered to be a mapping from a set of units \( \Omega \) in a set of distributions.

We consider a data table of N rows and p column where each row is associated to a unit belonging in the set \( \Omega \) and each column is defined by a variable such that the cell of this data table, associated to a given unit and a given variable is a distribution. This set of given initial distributions is called the "distribution base".

We consider first the case of a unique variable "Y" with domain \( Y \subseteq \Omega \) called description set. We denote \( X_i \) the random variable associated to the ith unit (or "individual") and \( F_i \) its associated distribution defined by : \( F_i(t) = \Pr(X_i \leq t) \). The distribution base is now reduced to the set \( F = \{ F_i / i = 1,N \} \).
First we define a "point-distribution of distributions" associated to the given variable $Y$ at a point $T_n$ by:

$$G_{T_n}(x) = \Pr\{F_i \in F / F_i(T_n) \leq x\} = \text{card}\{F_i \in F / F_i(T_n) \leq x\}/N,$$

where $x \in [-\infty, +\infty]$. For instance, if $x$ is the median (i.e. $x = 1/2$), $G_{T_n}(x)$ is the percentage of units whose probability of taking a value less than $T_n$ is less than $1/2$.

We define a "k-point joint distribution of distributions" by:

$$H_{T_1, \ldots, T_k}(x_1, \ldots, x_k) = \Pr\{F_i \in F / F_i(T_1) \leq x_1 \land \ldots \land F_i \in F / F_i(T_k) \leq x_k\}.$$

In the following, we suppose that $T_i$ is increasing with $i$ and $\text{Min}\{F_i(T_1)/F_i \in F\} > 0$. Also, in order to simplify notations, sometimes when there is no ambiguity, $H_{T_1, \ldots, T_k}(x_1, \ldots, x_k)$ is replaced by $H(x_1, \ldots, x_k)$.

**Proposition 1**

1. $G_{T_n}$ is a distribution.
2. $H_{T_1, \ldots, T_k}$ is a k-dimensional joint distribution function with margin $G_{T_1}, \ldots, G_{T_k}$

**Proof:** this results from the definition of a distribution function and a n-dimensional distribution function as $G_{T_n}$ is not decreasing and $G_{T_n}(-\infty) = 0$, $G_{T_n}(\infty) = 1$. Also, $H_{T_1, \ldots, T_k}$ is n-increasing and $H_{T_1, \ldots, T_k}(x_1, \ldots, x_k) = 0$ for all $X = (x_1, \ldots, x_k)$ in IR^k such that $x_i = 0$ for at least one $i \in \{1, \ldots, k\}$. This comes from the fact that $\{F_i(T_1)/F_i \in F\} > 0$ as the $F_i$ the $T_i$ are increasing and $T_1$ is chosen such that: $\text{Min}\{F_i(T_1)/F_i \in F\} > 0$. It is also easy to see that $H_{T_1, \ldots, T_k}(\infty, \ldots, \infty) = 1$ as $H_{T_1, \ldots, T_k}(1, \ldots, 1) = 1$ and $H_{T_1, \ldots, T_k}$ is increasing.

**Definition of a k-copula** (Schweizer and Sklar (1983), Nelsen (1998)):

A k-copula is a function $C$ from $[0, 1]^k$ to $[0, 1]$ with the following properties:

1. For every $u$ in $[0, 1]^n$, $C(u) = 0$ if at least one coordinate of $u$ is 0
2. If all coordinate of $u$ are 1 except $u^*$ then $C(u) = u^*$
3. The number assigned by $C$ to each hyper-cube $[a_1, a_2] \times [b_1, b_2] \times \ldots \times [z_k, z_k]$ is non negative.

For example, in two dimensions ($k = 2$), the third condition gives:

$$C(a_2, b_2) - C(a_2, b_1) - C(a_1, b_2) + C(a_1, b_1) \geq 0.$$ 

See figure 1:

or $a^*(w, \eta) = [Y(w) R(\eta) C(G^* T)]$.

---

**Figure 1:** the third condition of the definition of a copulas $C$ gives:

$$C(a_2, b_2) - C(a_2, b_1) - C(a_1, b_2) + C(a_1, b_1) \geq 0$$
Proposition 2
There exist a k-copula C such that for all \(X = (x_1, \ldots, x_k)\) \(\in [-\infty, +\infty]^k\):
\[
H_{T_1, \ldots, T_k}(x_1, \ldots, x_k) = C(G_{T_1}(x_1), \ldots, G_{T_k}(x_k)).
\]
Moreover, C is uniquely determined on \(\text{Ran} \; G_{T_1} \times \ldots \times \text{Ran} \; G_{T_k}\).

Proof: this results directly from Sklar theorem and proposition 1.

Parametric families of copulas
The most simple copulas denoted M, W and \(\Pi\) are \(M(u, v) = \min(u, v)\), \(\Pi(u, v) = u \cdot v\) and \(W(u, v) = \max(u+v-1, 0)\). These copulas are special cases of some parametric families of copulas as the followings:

- \(C_b(u, v) = \max([u^b + v^b -1]^{-1/b}, 0)\) discussed by Clayton (1978) has the following special cases: \(C_{-1} = W, C_0 = \Pi, C_\infty = M\).
- Frank (1979) has defined \(C_b(u, v) = -1/b \ln (1+(e^{-bu}-1)(e^{-bv}-1)/(e^{-b}-1))\) has the following special cases: \(C_{-\infty} = W, C_0 = \Pi, C_\infty = M\). Many other families are defined in Nelsen (1998).

In case of a Markov process \(X_t\) (of distribution \(G_t\) for instance) we have a transition formulas:
\[
C_{st} = C_{su} \ast C_{ut}
\]
where the product \(\ast\) is defined by: \((C_1 \ast C_2)(u,v) = \int_0^1 \partial C_1(u,t)/\partial u \cdot \partial C_2(t,v)/\partial v \; dt\). It is easy to verify that \(\Pi \ast C = C \ast \Pi\), \(M \ast C = C \ast M = C\) and \(W \ast W = M\). If the transition probability satisfies a standard Brownian motion we have:
\[
C_{st}(u,v) = \int_0^s \phi(\sqrt{(t-s)} \cdot \psi^{-1}(u) - \sqrt{(s-t)} \cdot \psi^{-1}(v)) \; dw
\]
where \(\phi\) denote the standard normal distribution function. For more details on the links between Markov processes and copulas, see Darsow, Nguyen, Olsen (1992).

Example: the distribution base (see figure 2) is reduced to two distributions \(F_1\) and \(F_2\).

![Figure 2: The distribution base is reduced to \(F_1\) and \(F_2\).](image)

![Figure 3: The copulas \(C(u, v)\), for instance \(C(0, 1) = 0, C(1,1/2) = 1/2.\)](image)
For instance:
\[
G_{T_1}(x_1) = \Pr(\{F_i \in F / F_i(T_1) \leq x_1 \}) = 1
\]
\[
G_{T_2}(x_2) = \Pr(\{F_i \in F / F_i(T_2) \leq x_2 \}) = 1/2
\]
\[
H_{T_1,T_2}(x_1, x_2) = C(G_{T_1}(x_1), G_{T_2}(x_2)) = C(1, 1/2) = 1/2
\]
From all calculation of \(C(u, v)\), (see figure 3), it results that \(C = \text{Min}\).

Example
The space is partitioned in cubes. A distribution of humidity is associated to each cube: \(F_i(t) = \Pr(X_i \leq t)\) where \(X_i(w)\) is the humidity at a point \(w\) which belongs in the cube \(i\). Hence, the distribution base is the set of all their associated distributions and \(F_i(t)\) is the proportion of points in the cube which humidity is lower than \(t\).
\(G_{T_n}(x)\) is the proportion of \(F_i\) such that \(F_i(T_n) \leq x\). For instance, if \(x\) is a quartile (i.e. \(x = 1/4\)), \(G_{T_n}(x)\) is the percentage of cubes whose probability of taking a humidity lower than \(T_n\) is less than \(1/4\).
If for any \(x\), \(G_{t1}(x) = 0\) and \(G_{t2}(x) = 1\), then this means that the humidity of the cube varies between \(t1\) and \(t2\).
\(H_{T_1,T_2}(x_1, x_2) = C(G_{T_1}(x_1), G_{T_2}(x_2))\) is the proportion of cubes which humidity distribution is lower than \(x_1\) and \(x_2\) respectively at \(T_1\) and \(T_2\) (i.e. the proportion of cube distributions which have a probability of humidity less than \(x_1\) at \(T_1\) and less than \(x_2\) at \(T_2\)).

Proposition 3
If each point distribution \(G_{T_1}, ..., G_{T_k}\) converges, when the size of the distribution base \(F\) grows to infinity, then the \(k\)-dimensional joint distribution function \(H\) with margin \(G_{T_1}, ..., G_{T_k}\) converges.

Proof
This come from the fact that a copula is continuous in its horizontal, vertical and diagonal section (see corollary 2.2.6 in Nelsen (1998)).

2. Symbolic objects associated to a distribution base
A concept is defined by an intent (its characteristic properties, also called its "description") and an extent (the units which "satisfy" these properties). Here, each unit is described by a set of distribution. Together the units define the distributions base and are supposed to satisfy the properties of a given concept. For instance, the units are towns described by socio-economic distributions (as the age or wages distribution of their inhabitant) and the concept is the region containing these towns. More formally, if \(C\) is a region, \(\text{Extent}(C)\) is the set of towns of this region \(\text{Intent}(C) = d_c\) is a description of the region. A symbolic object (see Diday (1998) or Bock, Diday (2000) for more details) is a model for a concept \(C\), it is defined by a triplet \(s = (a, R, d_c)\) where here
- \(d_c\) is the \(k\)-point joint distribution of distributions of the distribution base: \(H_{T_1, ..., T_k}\)
- \(R\) is a binary relation between descriptions such that the value \([d'Rd]\) \(\in [0,1]\) measures the degree to which \(d'\) is in relation with \(d\) (see Bandemer and Nather (1992)).
- "a" measures the "fit" between a unit "w" and the concept \(C\). It is a mapping from \(\Omega\) into \([0,1]\) such that \(a(w) = [Y(w) R d_c]\) where \(Y(w)\) is the \(k\)-point joint distribution of distributions of a distribution base reduced to \(Y(w)\).

The fit between \(w\) and \(C\) can be measured by different ways, for instance, by the "density of distributions" around \(Y(w)\) among the set of descriptions \(\{Y(w') / w' \in \text{Extent}(C)\}\). In this context, the "density" can be defined in two ways in using, first an approximation of the derivative of the copula (see hereunder 2.1), second the derivative of the copula, when it exists! (see 2.2). The fit between \(w\) and \(C\) can also be measured by comparing the \(k\)-point joint distribution of
distributions associated to the unit "w" and to the concept C (see 2.3). In order to simplify, we consider the case \( k = 2 \) where each unit is described by only two distributions. It is then easy to extend this case to a multivariate copula by using the H-volume (defined in Nelsen (1998), p. 36).

2.1 Fit between a unit and a concept by using an approximation of the density

We define a symbolic object \( s = (a, R(\eta), d) \) such that:

\[
a(w, \eta) = [Y(w) R(\eta) C_s(G_{T_1}, G_{T_2})] \in \mathbb{R}^+ \text{ which measures the "fit" between a distribution } L = Y(w) \text{ and the copula } C_s(G_{T_1}, G_{T_2}) = H_{T_1, T_2} \text{ associated to the distribution base } F.
\]

With \( x_i = L(T_i) \) (the value of the distribution \( L \), at the point \( T_i \)), \( R(\eta) \), with \( \eta = (\varepsilon_1, \varepsilon_1) \) is defined by:

\[
a(w, \eta) = [Y(w) R(\eta) C_s(G_{T_1}, G_{T_2})] = (C_s(G_{T_1}(x_1 + \varepsilon_1), G_{T_2}(x_2 + \varepsilon_2)) - C_s(G_{T_1}(x_1 + \varepsilon_1), G_{T_2}(x_2 - \varepsilon_2)) - C_s(G_{T_1}(x_1 - \varepsilon_1), G_{T_2}(x_2 + \varepsilon_2)) + C_s(G_{T_1}(x_1 - \varepsilon_1), G_{T_2}(x_2 - \varepsilon_2))).
\]

Hence, \( a(w, \eta) = [Y(w) R(\eta) C_s(G_{T_1}, G_{T_2})] \) with \( x_i = Y(T_i) \). Notice that due to proposition 1, \( G_{T_1} \) and \( G_{T_2} \) are increasing, so we have \( G_{T_i}(x_i + \varepsilon_i) \geq G_{T_i}(x_i - \varepsilon_i) \).

**Proposition 4**

We have \( a(w, \eta) = [Y(w) R(\eta) C_s(G_{T_1}, G_{T_2})] \in [0, 1] \).

**Proof:** This can be proved in the following way:

\[
a(w, \eta) = (H(x_1 + \varepsilon_1, x_2 + \varepsilon_2) - H(x_1 + \varepsilon_1, x_2 - \varepsilon_2) - H(x_1 - \varepsilon_1, x_2 + \varepsilon_2) + H(x_1 - \varepsilon_1, x_2 - \varepsilon_2)).
\]

2.2 Fit between a unit and a concept by using the derivative of a copula

Let \( G_T = (G_{T_1}, G_{T_2}) \) and \( H_{T_1,...,T_k} = C(G_T) \). When the derivative of \( G \) and of the copula \( C \) exist, we can calculate the derivative of a k-point joint distribution of distributions, in the following way.

The density function \( h \) associated to the distribution of distributions \( H \), is the derivative of \( C(G_T) \) at \( X = (x_1, x_2) \). It is given by:

\[
h(X) = H'_{T_1,T_2}(X) = \frac{\partial^2 H_{T_1,T_2}(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 C(G_{T_1}(x_1), G_{T_2}(x_2))}{\partial x_1 \partial x_2}
\]

\[
h(X) = \frac{\partial}{\partial x_1} \cdot \frac{\partial}{\partial x_2} H_{T_1,T_2}(x_1, x_2) = \frac{\partial^2 C(G_{T_1}(x_1), G_{T_2}(x_2))}{\partial x_1 \partial x_2}
\]

\[
h(X) = \frac{\partial}{\partial x_1} \cdot \frac{\partial}{\partial x_2} H_{T_1,T_2}(x_1, x_2) = \frac{\partial}{\partial x_1} \left[ \frac{\partial^2 C(G_{T_1}(x_1), G_{T_2}(x_2))}{\partial u_1 \partial u_2} \right]
\]

\[
h(X) = G_{T_1}'(X) C'(G_T(X)) \quad \text{where} \quad C'(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \text{ and} \quad G_{T_1}'(x_1, x_2) = \frac{\partial}{\partial x_1} G_{T_1}(x_1) \cdot \frac{\partial}{\partial x_1} G_{T_2}(x_2)
\]

From the proposition 4 it results:

**Proposition 5**

When the derivative of \( G \) and of the copula \( C \) exist, when \( \eta \rightarrow 0 \),

\[
a(w, \eta) = [Y(w) R(\eta) C(G_{T_1}, G_{T_2})] \rightarrow a(w,0) = h(L(T_1), L(T_2)) \in [0,1] \text{ where } Y(w) = L.
\]

**Proof**

This results from the definition of \( a(w, \eta) \) which converges by definition of a derivative towards the derivative \( h \) of \( H \) when \( \eta \rightarrow 0 \). Moreover, as from the proposition 4, \( a(w, \eta) \in [0,1] \), we have \( h(L(T_1), L(T_2)) \in [0,1] \).

2.3 Fit between a unit and a concept by comparing two joint distribution of distributions.
We call $H_{T_1, \ldots, T_k, w}$ the $k$-points joint distribution of distributions associated to the distribution base $F^* = \{ Y(w) \}$ reduced to the single distribution $Y(w)$ denoted $F_w$. In this case, we have the following result:

**Proposition 6**

The $k$-copula $C$ associated to the $k$-points joint distribution of distributions satisfies the following properties: i) its domain is the set $\{0, 1\}$, ii) $C = \text{Min}$ or $C = \prod$.

**Proof:**

By definition of a "$k$-point joint distribution of distributions":

$$H_{T_1, \ldots, T_k, w}(x_1, \ldots, x_k) = \Pr(\{F_w(T_1) \leq x_1\} \land \ldots \land \{F_w(T_k) \leq x_k\})$$

which is equal to 1 if $\forall i$ we have $F_w(T_i) \leq x_i$ and it is equal to 0 if $\exists i$ such that $F_w(T_i) > x_i$. In other words,

$$H_{T_1, \ldots, T_k, w}(x_1, \ldots, x_k)$$

is equal to 1 if $\forall i F_{T_i}(x_i) = 1$ and is equal to 0 if $\exists i$ such that $G_{T_i}(x_i) = 0$. It results that $H_{T_1, \ldots, T_k, w}(x_1, \ldots, x_k) = C(G_{T_1}(x_1), \ldots, G_{T_k}(x_k))$ belongs to $\{0, 1\}$ and moreover the copula $C$ is the product or the min which leads in this special case to the same result.

Let $G_T = (G_{T_1}, \ldots, G_{T_k})$ and $H(Y) = H_{T_1, \ldots, T_k, w}$. We define the following symbolic object associated to the distribution base $F^*$ and $F$ by:

$$s = (a, R, C(G_T))$$

with $a(w) = \{H(Y) R C(G_T)\} = \{H_{T_1, \ldots, T_k, w} R C(G_T)\}$ where $R$ measures the degree to which the $k$-points $(T_1, \ldots, T_k)$ joint distribution of distributions $H_{T_1, \ldots, T_k, w}$ associated to the unit $w$ is in relation with the $k$-points $(T_1, \ldots, T_k)$ joint distribution of distributions associated to the distribution base $F$ associated to a concept. Notice that the $R$ measure can be chosen among standard dissimilarities between distributions function extended to joint distributions function (as Paul Levy, Hellinger, Kullback-Leibler, see Tassy, Legait (1990) for a review).

### 2.4 Partitioning algorithms and criteria to optimize

The quality on a subset $A$ of $\Omega$ of the copula model $C$ when the model of $G$ is chosen can be given by: $Q(C, G) = \prod_{w \in A} a(w)$ or $Q(C, G) = \text{Min}_{w \in A} a(w)$ or $Q(C, G) = (\sum_{w \in A} a(w))/N$ to be maximised.

Hence, when we look for a partitioning, the criteria $Q$ has to be maximised for each class.

### 2.5 The choice of $T_j$

As we are looking for a partition of the set of distributions, it is sure that a given $T_j$ is bad if all the distribution $F_i$ of the base $F$ take the same value at $T_j$. Also $T_i$ is a bad choice if all the $F_i(T_j)$ are uniformly distributed in $[0, 1]$. In fact we can say that a $T_j$ is good if distinct classes of values exist among the set of values: $\{ F_i(T_j) / i = 1, N \}$. In Jain and Dubes (1988) several methods are proposed in order to reveal clustering tendency. Here we are in the special case were we look for such tendency among a set of points belonging in the interval $[0, 1]$.

We suggest a method based on the number of triangles whose vertices are points of $[0, 1]$ and where closest sides are larger (resp. smaller) than the remaining side and denoted $A$ (resp. $B$). We define the hypotheses $H^0$ that there is no clustering tendency, by the distribution of a random variable $X^0$ which associates to $N$ points randomly distributed in the interval $[0, 1]$ and denoted $u$, the value $X^0(u) = A - B / C^3_N = 6(A-B)/n(n-1)(n-2)$ which belongs in $[-1, 1]$. The greater is $X^0(u)$ the higher is the clustering tendency. We calculate the number of triangles whose vertices are points of $U = \{ F_i(T_j) / i = 1, N \}$ for which the two closest sides are larger (resp. smaller) than the remaining side and denoted $A(U)$ (resp. $B(U)$). Having the distribution of $X^0$, the value $A(U) - B(U) / C^3_N$ can reject or accept the null hypotheses at a given threshold.

### 3. The mixture decomposition of distributions of distributions in classification
3.1 The mixture decomposition problem

The problem can be settled in the following way:

Given k, find a partition \( P = (P_1, ..., P_k) \) of \( F \) and the value of k parameters \((\alpha_1, ..., \alpha_k)\), such that each class \( P_i \) be considered as a sub-sample which follows the k-dimensional joint distribution function \( H_i(., \alpha_i) \):

\[
H(X) = \sum_{i=1,k} p_i H_i(X, \alpha_i) \text{ where } p_k \in [0,1] \text{ and } \sum_{i=1,k} p_i = 1.
\]

In order to solve this problem we can settle it in term of mixture decomposition of density functions, by setting in case of \( H \) derivability: \( h(X) = H'(X) \) and \( h_i(X, \alpha_i) = H'_i(X, \alpha_k) \). Then the mixture decomposition problem becomes: given k, find a partition \( P = (P_1, ..., P_k) \) of \( F \) and the value of k parameters \((\alpha_1, ..., \alpha_k)\), such that each class \( P_i \) be considered as a sub-sample which follows the k-dimensional joint distribution function \( h_i(., \alpha_i) \):

\[
h(X) = \sum_{i=1,k} p_i h_i(X, \alpha_i).
\]

The parameters \( \alpha_i = (d_i, b_i) \) depends on the parameters of a chosen copulas family model (for instance, the Frank family, see 2.5) denoted \( b_i \) and on the parameters of a chosen distribution family model denoted \( d_i \) (defined on \([0,1]\) like a Dirichlet or Multinomial distribution).

In order to approximate the density function \( h_i(X, \alpha_i) \) or to calculate it, we can use the two ways presented in section 2. In the first case (see 2.1), which can be also used in case of not derivability of \( H \) we settle:

\[
h_i(X, \alpha_i, \eta_i) = a_i(w, \eta_i) = [Y(w) R(\eta_i) C_i (G^{T_1}_1(d_{i1}), G^{T_2}_1(d_{i2}), b_i)] \text{ where } X = Y(w), \alpha_i = (d_i, b_i) \text{ with } d_i = (d_{i1}, d_{i2}) \text{ and } \eta_i = (\epsilon_{1i}, \epsilon_{2i}).
\]

Hence, the mixture decomposition model can be settled in the following way:

\[
h(X) = \sum_{i=1,k} p_i h(X, \alpha_i, \eta_i).
\]

In the second case (see 2.2), where \( H, C \) and \( G \) are supposed derivable, we settle:

\[
h_i(X, \alpha_i) = C_i(G(X, d_i), b_i) \text{ with } \alpha_i = (d_i, b_i) \text{ and } C_i = C_i'(G(X, d_i), b_i) = \sum_{i=1,k} G_i(X, d_i) C_i'(G(X, d_i), b_i) \text{ where } \]

\[
C_i' = \frac{\partial C_i}{\partial u_1} \frac{\partial u_1}{\partial x_1} + \frac{\partial C_i}{\partial u_2} \frac{\partial u_2}{\partial x_2}.
\]

In the following, we denote \( \beta_i = (\alpha_i, \eta_i) \) in the first case and \( \beta_i = \alpha_i \) in the second case . Hence, the general decomposition model can be expressed in the following way:

\[
h(X) = \sum_{i=1,k} p_i h(X, \beta_i).
\]

3.2 The symbolic mixture decomposition problem and four algorithms for solving it:

Given the models associated to \( G \) and \( C \), the decomposition can be obtained by maximizing alternatively (Diday et al 1974) a criterion. With \( L = Y(w), x_i = L(T_i) \) and \( X = (x_1, x_2) \) this criterion can be, the likelihood :

\[
v(X, \beta) = \prod_{i=1,k} p_i \Pi_{w \in P_i} h_i(X, \beta_i)) \text{ where } \beta = (\beta_1, ..., \beta_k) \text{ are the parameters of the densities } h_i,
\]

the log likelihood:

\[
V(X, \tilde{\beta}) = \sum_{i=1,k} \sum_{w \in P_i} \log (p_i h_i(X, \beta_i)).\]

Or any measure of the fit between the sample and the distributions of the mixture decomposition as for instance:

\[
Z(X, \beta) = \sum_{i=1,k} \sum_{w \in P_i} h_i(X, \beta_i) \text{ which will be used in the following example given in 4.4.}
\]

We suggest the four following algorithms based on two steps of "maximisation" and "representation" in the framework of the so called "Nuées dynamique" or "dynamical clustering" method (Diday (1971), Diday and al (1979) where here the representation step is an estimation of the parameters of a copula model :

**Algorithm 1:**

Input: a set of units described by distributions and a number of classes.

Output: a partition and a copula model \( C_i \) for each class and optionally a distribution \( G_i \) for each class at each \( T_i \).

This algorithm is defined in two steps of representation and allocation Diday, Ok, Schroeder (1974):
1. **Representation:** \( g(P_1, \ldots, P_k) = (\beta_1, \ldots, \beta_k) \) by finding the parameters \( (\beta_1, \ldots, \beta_k) \) which maximise the chosen criterion (for instance \( v, V \) or \( Z \)).

2. **Maximisation:** \( f(\beta_1, \ldots, \beta_k) = (P_1, \ldots, P_k) \) by finding the partition \( (P_1, \ldots, P_k) \) where \( P_i \) is the set of distributions:\n\[ P_i = \{ X / p_i h_i (X, \beta_i) \geq p_m h_m (X, \beta_m) \} \] with \( i < m \) in case of equality}. When the criterion are bounded (which is the case of \( v, V, Z \)), it is easy to show that this algorithm converges as it increases the criterion at each step.

There are several variants for the choice of \( p_i \): \( p_i = \text{card } P_i / N \) at the last step, or at each step, see for instance, Celeux, Govaert (1993).

In the first case (defined in 2.1) where \( \eta_i = (\epsilon_{1i}, \epsilon_{2i}), \epsilon_i = \text{Min} (\epsilon_{1i}, \epsilon_{2i}) \) is increased until \( \text{Max}_i a_i (w, \alpha_i) \) becomes different from all (or a given number) of \( a_j (w, \alpha_j) \).  

**Algorithm 2:** SYEM

**Input:** a set of units described by distributions and a given number of classes.  
**Output:** \( k \) copula model \( C_i \) and optionally \( k \) associated distributions \( G_i \) for each class at each \( T_i \).  
This algorithm is based on the two steps of the EM algorithm (Dempster, Laird, Rubin (1977)):

- **Estimation:** for \( l = 1, k \) and any \( w \in \Omega \), \( t^n_l(X) = p^n_l h_l (X, \beta^n_l) / \sum_{w \in \Omega} p^n_l h_i (X, \beta^n_l) \) which is the posterior conditional probability of an individual \( w \) (with \( X = Y(w) \)) to belong to the class \( l \) at the \( n \)th iteration.
- **Maximisation:** with \( p^{n+1}_l = 1/N \sum_{w \in \Omega} t^n_l(Y(w)) \) maximise:
\[ \sum_{w \in \Omega} t^n_l(Y(w)) \partial \log h(Y(w), \beta^{n+1}_l) / \partial \beta^{n+1}_l \]

**Algorithm 3:** SYSEM

**Input:** a set of units described by distributions and a number of classes.  
**Output:** a family of admissible partitions, a copula model \( C_i \) for each class and optionally a distribution \( G_i \) for each class at each \( T_i \).  
This algorithm is based on the steps defined by the SEM algorithm (Celeux, Diebolt) which add a stochastic step to the EM algorithm  

**Algorithm 4**

Here we come back to the problem (see 4.1) of finding directly the joints \( H_i \) such that:
\[ H(X) = \sum_{i=1,k} p_i H_i (X, \alpha_i) \] where \( p_k \in [0,1] \) and \( \sum_{i=1,k} p_i = 1 \). To do so, we can use the symbolic object defined in 2.3. The decomposition problem can then be settled in the following way:

\[ [H_{T1, \ldots, T_k, w} R H] = \sum_{i=1,k} p_i [H_{T1, \ldots, T_k, w} R H_i] \] which can be written

\[ [H_{T1, \ldots, T_k, w} R C(G_{T})] = \sum_{i=1,k} p_i [H_{T1, \ldots, T_k, w} R C_i (G_{T}, \alpha_i)] \]

1. **g(P_1, \ldots, P_k) = (\alpha_1, \ldots, \alpha_k) by maximization of the chosen criterion (induced from \( v, V \) or \( Z \))**

2. **f(\alpha_1, \ldots, \alpha_k) = (P_1, \ldots, P_k) where P_i is the set of distributions:**
\[ P_i = \{ X / p_i h_i (X, \beta_i) \geq p_m h_m (X, \beta_m) \} \] with \( i < m \) in case of equality}.

### 3.3 A simple example of mixture decomposition

The symbolic data table is defined by five units \( w_i \), each one described by a distribution \( F_i \). The five distributions are given on figure 4.
The family of copulas parametric model is simple: \( b_i \in \{0,1\} \), \( C(u,v) = M \) if \( b_i = 1 \) and \( C(u,v) = W \) (see 2.4) if \( b_i = 0 \).

There is no parametric model for \( G_{T_1} \) and \( G_{T_2} \) as we use only their empirical distribution. Hence, \( G_{T_j}(x) = \Pr \{ L \in P_i / L(T_j) \leq x \} \) and there are no parameters \( d_i \).

In order to simplify calculations, we use the following criterion:

\[
Z(X, \beta) = \sum_{i=1,2} p_i \sum_{w \in P_i} h_i (X, \beta_i) = \sum_{i=1,2} p_i \sum_{w \in P_i} C((G_{T_1}(x_1+\epsilon_{1i}), G_{T_2}(x_2+\epsilon_{2i})), b_i) - C((G_{T_1}(x_1+\epsilon_1), G_{T_2}(x_2-\epsilon_2)), b_i) - C((G_{T_1}(x_1-\epsilon_1)), G_{T_2}(x_2+\epsilon_2)), b_i) + C((G_{T_1}(x_1-\epsilon_1), G_{T_2}(x_2-\epsilon_2)), b_i).
\]

The initial given partition is \( P = (P_1, P_2) \) with \( P_0 = \{ F_1, F_2, F_3 \} \) and \( P_0 = \{ F_4, F_5 \} \).

We choose \( \epsilon_i = (\text{Max}_{L \in F} G_{T_i}(L(T_i)) - \text{Min}_{L \in F} G_{T_i}(L(T_i)))/N \). So we find \( \epsilon_1 = \epsilon_2 \approx 0.1 \).

First step: induction of parameters from classes \( g(P_1,...,P_k) = (\beta_1,...,\beta_k) \) by maximisation of the criterion \( Z \) where \( X_i = (x_{i1}, x_{i2}) \), \( x_i = L_i(T_i) \) and \( L_i = \text{Y}(w_i) \).

Here we have: \( \beta_i = (\alpha_i, \eta_i) \) where \( \alpha_i \) is generally equal to \( (d_i, b_i) \) is reduced to \( b_i = (b_1, b_2) \) and \( \eta_i = (\epsilon_{1i}, \epsilon_{2i}) = (0.1, 0.1) \). For each \( w \) \( \in \{ w_1, ..., w_5 \} \) and for \( b_{ij} \in \{0,1\} \) (i.e. \( C \in \{ \text{Min},W \} \)), we compute:

\[
e^i_1 = C^i((G_{T_1}(x_1+\epsilon_1), G_{T_2}(x_2+\epsilon_2)), bij), e^i_2 = - C^i((G_{T_1}(x_1+\epsilon_1), G_{T_2}(x_2-\epsilon_2)), b), e^i_3 = - C^i ((G_{T_1}(x_1-\epsilon_1)), G_{T_2}(x_2+\epsilon_2)), b), e^i_4 = C^i ((G_{T_1}(x_1-\epsilon_1), G_{T_2}(x_2-\epsilon_2)), b).\]

Hence, we obtain the following table:

<table>
<thead>
<tr>
<th>class Copula Individual : w1</th>
<th>Individual : w2</th>
<th>Individual : w3</th>
<th>( \Sigma_{w \in P_i} h_i (X, \beta_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_1 = { w_1, w_2, w_3 } )</td>
<td>( e^{i1}_1, e^{i2}_2, e^{i3}_3, e^{i4}_4(w2) )</td>
<td>( e^{i1}_1, e^{i2}_2, e^{i3}_3, e^{i4}_4(w2) )</td>
<td>( \sum_{i=1,4} e^{i(w_1)} )</td>
</tr>
<tr>
<td>W</td>
<td>1/3, 0, 0, 0</td>
<td>1/3, 0, 0, 0</td>
<td>1, -2/3, -2/3, 1/3</td>
</tr>
<tr>
<td>Min</td>
<td>2/3, -1/3, 0, 0</td>
<td>2/3, 0, -1/3, 0</td>
<td>1, -2/3, -2/3, 2/3</td>
</tr>
<tr>
<td>class Copula Individual : w4</td>
<td>Individual : w5</td>
<td>( \Sigma_{w \in P_i} h(X, \beta_i) )</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>( P_2 = { w_4, w_5 } )</td>
<td>( e^{i1}_1, e^{i2}_2, e^{i3}_3, e^{i4}_4 )</td>
<td>( e^{i1}_1, e^{i2}_2, e^{i3}_3, e^{i4}_4 )</td>
<td>( \sum_{i=1,4} e^{i(w_1)} )</td>
</tr>
<tr>
<td>W</td>
<td>1, 0, 0, 0</td>
<td>1, -1/2, -1/2, 0</td>
<td>1</td>
</tr>
<tr>
<td>Min</td>
<td>1, 0, 0, 0</td>
<td>1, -1/2, -1/2, 1/2</td>
<td>3/2</td>
</tr>
</tbody>
</table>
Table 1: Induction of the copulas from the classes $P^0_1$ and $P^0_2$.

As $p_1 = 3/5$ and $p_2 = 2/5$, it results from the Table 1, that for $b_1 = b_2 = \text{Min}$, $Z(X, \beta) = 6/5$, for $b_1 = b_2 = W$, $Z(X, \beta) = 4/5$, for $b_1 = \text{Min}, b_2 = W$, or $b_1 = W, b_2 = \text{Min}$, $Z(X, \beta) = 1$. Therefore $b_1 = b_2 = \text{Min}$ gives the best solution. In other words the best parameters are defined by $b_1 = 1$ for the class $P^0_1$ and $b_2 = 1$ for the class $P^0_2$.

Step 2: for each individual $w$ we compute $h_i(X, \beta_i) = a_i(w, \eta) = \left[ Y(w) R(\eta) C^i((G^iT_1, G^iT_2)) \right]$ and the class membership is $i$ when: $a_i(w, \eta)/\text{card}(P^0_i) > a_j(w, \eta)/\text{card}(P^0_j)$.

<table>
<thead>
<tr>
<th>Individual</th>
<th>$e^i_1, e^i_2, e^i_3, e^i_4$</th>
<th>$a_1(w, \eta)$</th>
<th>$e^{\bar{i}}_1, e^{\bar{i}}_2, e^{\bar{i}}_3, e^{\bar{i}}_4$</th>
<th>$a_2(w, \eta)$</th>
<th>Class membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>1/3</td>
<td>0, 0, 0, 0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>w2</td>
<td>1/3</td>
<td>0, 0, 0, 0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>w3</td>
<td>1/3</td>
<td>1/2, 0, 0, 0</td>
<td>1/2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>w4</td>
<td>1, 1, 1, 1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>w5</td>
<td>1, 1, 1, 1</td>
<td>0</td>
<td>1/2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: allocation of each individual to the best fit class. Some $e^i_1$ are already computed in table 1.

It results that the new classes are $P^1_1 = \{F1, F2\}$ and $P^1_2 = \{F3, F4, F5\}$.

Step 3: induction of parameters from classes

<table>
<thead>
<tr>
<th>class</th>
<th>Copula</th>
<th>Individual : w1</th>
<th>Individual : w2</th>
<th>$\sum_{w \in P^1_1} h_1(X, \beta_1)$</th>
<th>$\sum_{i=1,4j=1,3} e^i_j(w_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^1_1$</td>
<td>${w_1, w_2, w_3}$</td>
<td>$e^1_1, e^1_2, e^1_3, e^1_4$</td>
<td>$e^{\bar{1}}_1, e^{\bar{1}}_2, e^{\bar{1}}_3, e^{\bar{1}}_4$</td>
<td>(w2)</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>1, 0, 1/2, 0</td>
<td>1, 1/2, 0, 0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>1, -1/2, 0, 0</td>
<td>1, 0, -1/2, 0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>class</td>
<td>Copula</td>
<td>Individual : w3</td>
<td>Individual : w4</td>
<td>Individual : w5</td>
<td>$\sum_{w \in P^1_2} h(X, \beta_1)$</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>$P^1_2$</td>
<td>${w_4, w_5}$</td>
<td>$e^2_1, e^2_2, e^2_3, e^2_4$</td>
<td>$e^{\bar{2}}_1, e^{\bar{2}}_2, e^{\bar{2}}_3, e^{\bar{2}}_4$</td>
<td>$e^{\bar{2}}_1, e^{\bar{2}}_2, e^{\bar{2}}_3, e^{\bar{2}}_4$</td>
<td>$\sum_{i=1,4j=4,5} e^i_j(w_j)$</td>
</tr>
<tr>
<td>W</td>
<td>1/3, 0, 0, 0</td>
<td>1, -1/3, -1/3, 0</td>
<td>1, -2/3, -2/3, 1/3</td>
<td>2/3</td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>2/3, 0, 0, 0</td>
<td>1, -1/3, -1/3, 1/3</td>
<td>1, -2/3, -2/3, 1/3</td>
<td>2/3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Induction of the copulas from the classes $P^1_1$ and $P^1_2$. It results that $W$ for the first class and $\text{Min}$ for the second class are the best.

Step 4: Building new classes

<table>
<thead>
<tr>
<th>Individual</th>
<th>$e^i_1, e^i_2, e^i_3, e^i_4$</th>
<th>$a_1(w, \eta)$</th>
<th>$e^{\bar{i}}_1, e^{\bar{i}}_2, e^{\bar{i}}_3, e^{\bar{i}}_4$</th>
<th>$a_2(w, \eta)$</th>
<th>Class membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>1/2</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w2</td>
<td>1/2</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w3</td>
<td>0</td>
<td>2/3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w4</td>
<td>0</td>
<td>2/3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w5</td>
<td>0</td>
<td>1/3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: allocation of each individual to the best fit class. The new partition $(P^1_1, P^1_2)$ is the same as the preceding: $(P^0_1, P^0_2)$.  

The process has converged towards the partition: \((P^1_1, P^1_2) = \{\{F1, F2\}, \{F3, F4, F5\}\} and the copulas \(W\) for the first class and \(\text{Min}\) for the second.

4) The special case of the standard mixture decomposition problem

4.1 Properties of a distribution base of unit mass distributions

Imbedding the standard mixture decomposition problem in the mixture decomposition of distribution of distributions problem, in the case of a unique quantitative random variable \(Z\), our aim in this section. Each value taken by an individual "w" can be transformed in a unique way in a distribution which takes the value 0 until \(Z(w)\) not included and the value 1 after. Such distribution is called "unit mass". More formally, if \(\Omega = \{w_1, \ldots, w_n\}\) and \(Z(w_i) = z_i\), the distribution \(F_i\) associated to \(w_i\) is defined by \(F_i(t) = \text{Pr}(X_i \leq t)\) where the random variable \(X_i\) associated to \(w_i\) is such that its distribution \(F_i\) satisfies: \(F_i(t) = 0\) if \(t < z_i\) and \(F_i(t) = 1\) if \(t \geq z_i\). If the distribution base \(F\) contains only such distributions \(F_i, \ F_z\ is the distribution associated to the random variable \(Z\) and the \(T_i\) increases with \(i\), we have the following results:

**Proposition 7**

1) If \(H(x_1, \ldots, x_p) = C(G_{T_1}(x_1), \ldots, G_{T_p}(x_p))\) then \(C\) is the \(\text{Min}\) copulas.

2) If \(x_p < 1\), then \(\text{Min}(G_{T_1}(x_1), \ldots, G_{T_p}(x_p)) = G_{T_p}(x_p)\).

3) If \(x < 1\) then \(G_{T}(x) = \text{Prob}(Z > T) = 1 - F_z(T)\).

4) If \(x_p < 1\) then \(F_z(T_p) = 1 - H(x_1, \ldots, x_p)\).

**Proof**

**Lemma**

If \(F\) is a set of unit mass distributions, \(x_i \in [0, 1]\) for \(i = 1, \ldots, j\), \(A_j = \{f \in F/ f(T_j) = 0\}\) and \(B_j = \{f \in F/ f(T_j) \leq x_j, 1 \leq i \leq j\}\), then we have: \(A_j = B_j\) and \(|A_j| = \text{Min}_{i=1,j} |A_i|\).

**Proof**

\(B_j \subseteq A_j\) as \(f \in B_j\) implies \(f \in F\) and \(f(T_j) < x_j\) by definition of \(B_j\). As \(F\) is a set of unit mass distributions and \(x_j \in [0, 1]\), we have necessarily \(f(T_j) = 0\). Therefore \(f \in A_j\).

\(A_j \subseteq B_j\) as \(f \in A_j\) implies \(f(T_j) = 0\) which implies \(f(T_i) = 0\) for \(i = 1, \ldots, j\) as \(f\) is increasing as it is a distribution. So \(f \in B_j\) and therefore \(A_j = B_j\). As by definition of \(B_i\) we have \(B_i = \cap_{i=1,j} A_i\) and moreover we have proved that \(A_j \subseteq B_j\), it results that \(|A_j| = \text{Min}_{i=1,j} |A_i|\).

With this lemma, we can now prove the proposition 7.

1) If \(H(x_1, \ldots, x_p) = C(G_{T_1}(x_1), \ldots, G_{T_p}(x_p))\) then \(C\) is the \(\text{Min}\) copulas.

This can be proved in the following way: if all \(x_i\) are equal to 1, as all the elements of a distribution base take a value smaller than 1 everywhere, we have by definition of a point distribution of distributions: \(G_{T_i}(x_i) = 1\) and also, by definition of a \(k\)-point joint distribution of distribution \(H(x_1, \ldots, x_p) = 1\). So, in that case 1) is true. Suppose now that some \(x_i\) are smaller than 1 and suppose by denoting them \(x'_1, \ldots, x'_j\) such that their corresponding \(T\) denoted \(T'_1, \ldots, T'_j\) are increasing, then we have \(H(x'_1, \ldots, x'_j) = H(x_1, \ldots, x_p)\). This comes from the fact that the set of
distributions included in the distribution base, which are lower than $x'_i, \ldots, x'_j$ are the same as the ones which are also lower than $x_1, \ldots, x_p$. We can now apply the lemma 1 by denoting $A_j = \{ f \in F / f(T'_j) = 0 \}$ and $B_j = \{ f \in F / f(T'_j) \leq x'_i, 1 \leq i \leq j \}$. As $G_{T'_j}(x'_j) = A_j \cap F$ and $H(x'_i, \ldots, x'_j) = A_j \cap B_j$, it results that $G_{T'_j}(x'_j) = A_j \cap F$ and $H(x'_i, \ldots, x'_j) = B_j \cap F$. As from the lemma we have $A_j = B_j$ it results that $G_{T'_j}(x'_j) = H(x'_i, \ldots, x'_j)$ and so, $G_{T'_j}(x'_j) = H(x_1, \ldots, x_p)$. From the lemma, we have also $A_j = \min_{1 \leq i \leq j} A_i$ which implies, $G_{T'_j}(x'_j) = \min_{1 \leq i \leq j} G_{T_i}(x'_i)$. As $\min_{1 \leq i \leq j} G_{T_i}(x'_i) = \min_{1 \leq i \leq p} G_{T_i}(x_i)$ (as for the i such that $x_i = 1$ we have $G_{T_i}(x_i) = 1$ and for the i such that $x_i < 1$ we have $G_{T_i}(x_i) \leq 1$), we get $H(x_1, \ldots, x_p) = \min_{1 \leq i \leq p} G_{T_i}(x_i)$ which shows that $H(x_1, \ldots, x_p) = C(G_{T_1}(x_1), \ldots, G_{T_p}(x_p))$ where $C$ is the Min copulas.

2) If $x_p < 1$, then $\min (G_{T_1}(x_1), \ldots, G_{T_p}(x_p)) = G_{T_p}(x_p)$.

As in the proof of 1) we denote $x'_1, \ldots, x'_j$ (associated to increasing $T'_1, \ldots, T'_j$), the $x_i$ among $x_1, \ldots, x_p$ which are strictly lower than 1. It results that $x'_j = x_p$ and so from lemma 1 that $\min (G_{T_1}(x'_1), \ldots, G_{T_p}(x'_j)) = G_{T_p}(x_p)$. We have

$\min (G_{T_1}(x_1), \ldots, G_{T_p}(x_p)) = \min (G_{T_1}(x'_1), \ldots, G_{T_j}(x'_j))$ as shown in the preceding proof.

Therefore we get finally: $\min (G_{T_1}(x_1), \ldots, G_{T_p}(x_p)) = G_{T_p}(x_p)$.

3) If $x < 1$ then $G_T(x) = \mathbb{P} (Z > T) = 1-F_z(T)$ as by definition $F_z(T) = \mathbb{P} (\{ Z(w) \leq T \})$ and $G_T(x) = \mathbb{P} (\{ F_i \in F / F_i(T) \leq x \})$ is exactly the proportion of unit mass distributions $F_i$ which value is $F_i(t) = 1$ only strictly after $T$ (i.e. $t > T$), as $F_i(t) = 0$ if $t < Z(w_i)$ and $F_i(t) = 1$ if $t \geq Z(w_i)$. In other words, this means that $G_T(x)$ is the proportion of individuals $w$ such that $Z(w) > T$.

4) If $x_p < 1$ then $F_{T_p}(x_p) = 1- H(x_1, \ldots, x_p)$ as from 1), $H(x_1, \ldots, x_p) = \min(G_{T_1}(x_1), \ldots, G_{T_p}(x_p))$ and from 2), $\min(G_{T_1}(x_1), \ldots, G_{T_p}(x_p)) = G_{T_p}(x_p)$ and from 3) $F_{T_p}(x_p) = 1- G_{T_p}(x)$.

4.2 The standard mixture decomposition problem is a special case

Here we need to introduce the following notations: $F_{zi}$ is the distribution associated to a quantitative random variable $Z_i$ defined on $\Omega = \{ w_1, \ldots, w_n \}$. $F^i$ is a distribution base whose elements are the unit mass distributions associated to each value $Z_i(w_i)$ (i.e. they take the value 0 for $t \leq Z_i(w_i)$ and 0 for $t < Z_i(w_i)$). $G^i_T$ is a point-distribution of distributions at point $T$ associated to the distribution base $F^i$. $H^i_{T_1, \ldots, T_k}$ is a k-point joint distribution of distributions associated to the same distribution base.

**Proposition 8**

If $H_{T_1, \ldots, T_k} = \sum_{i=1,p} p_i H^i_{T_1, \ldots, T_k}$ with $\sum_{i=1,p} p_i = 1$, then $F_z = \sum_{i=1,p} p_i F_{zi}$.

**Proof**

From the proposition 2, we have $H^i_{T_1, \ldots, T_k}(x_1, \ldots, x_p) = C^i (G^i_{T_1}(x_1), \ldots, G^i_{T_p}(x_p))$ where $C^i$ is a p-copula. Therefore $H_{T_1, \ldots, T_k}(x_1, \ldots, x_p) = \sum_{i=1,p} p_i C^i (G^i_{T_1}(x_1), \ldots, G^i_{T_p}(x_p))$.

We choose $x_p < 1$ and we use 1), 2), 3), 4) in proposition 7.

From 1) we get: $H_{T_1, \ldots, T_k}(x_1, \ldots, x_p) = \sum_{i=1,p} p_i \min (G^i_{T_1}(x_1), \ldots, G^i_{T_p}(x_p))$.

From 2) we get: $H_{T_1, \ldots, T_k}(x_1, \ldots, x_p) = \sum_{i=1,p} p_i G^i_{T_p}(x_p)$. 


From 3) we get: \[ H_{T_1,\ldots,T_k}(x_1,\ldots,x_p) = \sum_{i=1,p} p_i (1-H(T_p)) = 1-\sum_{i=1,p} p_i F_{z_i}(T_p). \]

From 4) we have \( F_z(T_p) = 1- H(x_1,\ldots,x_p) \) and therefore \( F_z(T_p) = \sum_{i=1,p} p_i F_{z_i}(T_p) \).

As the same reasoning can be done for any sequence \( T_1,\ldots,T_p \), it results finally that: \( F_z = \sum_{i=1,p} p_i F_{z_i} \).

### 4.3 Links between the generalised mixture decomposition problem and the standard one

It results from the proposition 8 that by solving the mixture decomposition of distribution of distributions problem we solve the standard mixture decomposition problem. This results from the fact that it is possible to induce \( F_{z_i}(T_1),\ldots,F_{z_i}(T_p) \), from \( G^i_{T_1}(x_1) \), \ldots, \( G^i_{T_p}(x_p) \) and therefore, the parameters of the chosen model of the density law associated to each \( Z_i \). Moreover, by choosing the "best model" among a given family of possible models (Gaussian, Gamma, Poisson,\ldots) for each \( Z_i \), we can obtain a different model for each law of the mixture. "Best model" means the Gaussian, Gamma or Poisson,\ldots model which fit the best with \( F_{z_i}(T_1),\ldots,F_{z_i}(T_p) \) for each \( i \). It would be interesting to compare the result of both approach: the mixture decomposition of a distribution of distributions algorithms, and the standard mixture distribution algorithms in the standard framework, when the same model is used for each class or more generally when each law of the mixture follows a different family model.

### Conclusion

Many things remain to be done. For instance, to compare the results obtained by the general methods and the standard methods of mixture decomposition on standard data. We have considered the mixture decomposition problem in the case of a unique variable. In order to extend it to the case of several variables, we can proceed as follows: we look for the variable which gives the best mixture decomposition criteria value in two classes and we repeat the process to each class thus obtained until the size of the classes becomes to small. To extend the methods to the case where each class may be modelled by a different copula family. Also, the \( G_T \) can be modelled at each \( T \) by a different distribution family. We can also add other criteria taking care of a class variable and a learning set. Notice that the same approach can be used in the case where instead of having distributions we have any kind of mapping. For example, if each unit is described by a trajectory of a stochastic random process \( X_T \), then each \( G_T \) is the distribution of \( X_T \). We can then apply the same general approach in order to obtain classes of units characterised by copulas. For instance, in case of a Markov process, each class can satisfy a copula \( C_{st} = C_{su}*C_{ut} \). Instead of ditributions \( G_{Tn}(x) \) with \( T_n \) and \( x \in \mathbb{IR} \) we can generalise to \( T_n \) and \( x \in \mathbb{IR}^n \).

### References


