Design of Iterative Proportional Fitting Procedure for Possibility Distributions*

JIŘINA VEJNAROVÁ
Laboratory for Intelligent Systems, Prague, Czech Republic

Abstract

We design an iterative proportional fitting procedure (parameterized by a continuous t-norm) for computation of multidimensional possibility distributions from its marginals, and discuss its basic properties.

Keywords

multidimensional possibility distributions, marginal problem, triangular norm, iterative proportional fitting procedure

1 Introduction

The complexity of practical problems that are of primary interest in the field of artificial intelligence usually results in the necessity to construct models with the aid of a great number of variables: more precisely, hundreds or thousands rather than tens. However, distributions of such dimensionality are usually not available; the global knowledge (joint distribution) must be integrated on the basis of its local pieces (marginal distributions). This problem type is often referred to as a marginal problem. More precisely, the marginal problem addresses the question of whether or not a common extension exists for a given set of marginal distributions.

In [14] we introduced a possibilistic marginal problem and found necessary and sufficient conditions, respectively. This contribution is a natural continuation of our work — it tries to solve a practical problem: how to compute the values of an extension. Its aim is to introduce a possibilistic version of Iterative Proportional Fitting Procedure and to discuss its basic properties.

Iterative Proportional Fitting Procedure (IPFP) was originally designed by Deming and Stephan [3] in 1940 for adjustment of frequencies in contingency tables. Later, IPFP was applied to several problems in different domains; e.g. for

*This work has been partially supported by the GAČR Grant No. 201/02/1269 and the GA AV ČR Grant No. A1075104.

maximum likelihood estimate in a hierarchical model, or for computation of values of joint probability distributions in a probabilistic expert system [6] (for other applications see [9]).

This contribution is organized as follows. First an overview, followed by the basic notions (Section 2); then in Section 3 we briefly recall a possibilistic marginal problem, introduce possibilistic IPFP and demonstrate, on a simple example, how its computations are performed. In Section 4 we find a sufficient condition for its convergence and present two counterexamples.

2 Basic Notions

The purpose of this section is to give, as briefly as possible, an overview of basic notions of De Cooman's measure-theoretical approach to possibility theory [2], necessary for understanding the paper. We will start with the notion of a triangular norm, since most notions in this paper are parameterized by it.

2.1 Triangular Norms

A *triangular norm* (or a *t-norm*) T is an isotonic, associative and commutative binary operator on $[0,1]$ (i.e. $T:[0,1]^2 \to [0,1]$) satisfying the boundary condition: for any $x \in [0,1]$

$$T(1,x) = x.$$

Let $x,y \in [0,1]$ and T be a t-norm. We will call an element $z \in [0,1]$ *T-inverse* of x w.r.t. y if

$$T(z,x) = T(x,z) = y. \tag{1}$$

It is obvious that if $x \leq y$ then there are no T-inverses of x w.r.t. y. The *T-residual* $y \triangle_T x$ of y by x is defined as

$$y \triangle_T x = \sup\{z \in [0,1] : T(z,x) \leq y\}.$$

A t-norm T is called *continuous* if T is a continuous function. Within this paper, we will only deal with continuous t-norms, since for continuous t-norms $y \triangle_T x$ is the greatest solution of the equation (1) in z (if it exists).

Example 1 The most important examples of continuous t-norms are:

(i) *Gödel's t-norm:* $T_G(x,y) = \min(x,y)$;

(ii) *product t-norm:* $T_\Pi(x,y) = x \cdot y$;

(iii) *Lukasziewicz's t-norm:* $T_L(x,y) = \max(0, x+y-1)$;

and the corresponding residuals for $x > y$ (otherwise $y \triangle_T x = 1$ for any t-norm):

(i) $y \triangle_{T_G} x = y$;

(ii) $y \triangle_{T_\Pi} x = \frac{y}{x}$;
(iii) $y \triangle_{T_L} x = y - x + 1$.

Because of its associativity, any t-norm T can be extended to an n-ary operator $T^n : [0,1]^n \to [0,1]$, namely in the following way

$$\begin{aligned} T^2(a_1, a_2) &= T(a_1, a_2), \\ T^n(a_1, \ldots, a_n) &= T(T^{n-1}(a_1, \ldots a_{n-1}), a_n), \end{aligned}$$

for $n \geq 3$.

2.2 Possibility Measures and Distributions

Let $\mathbf{X}$ be a finite set called *universe of discourse* which is supposed to contain at least two elements. A *possibility measure* Π is a mapping from the power set $\mathcal{P}(\mathbf{X})$ of $\mathbf{X}$ to the real unit interval $[0,1]$ satisfying the following requirement: for any family $\{A_j, j \in J\}$ of elements of $\mathcal{P}(\mathbf{X})$

$$\Pi(\bigcup_{j \in J} A_j) = \max_{j \in J} \Pi(A_j)^1.$$

Π is called *normal* if $\Pi(\mathbf{X}) = 1$. Within this paper we will always assume that Π is normal.

For any Π there exists a mapping $\pi : \mathbf{X} \to [0,1]$, called a *distribution* of Π, such that for any $A \in \mathcal{P}(\mathbf{X})$, $\Pi(A) = \max_{x \in A} \pi(x)$. This function is a possibilistic counterpart of a density function in probability theory. In the remaining part of this contribution we will deal with distributions rather than with measures.

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ denote two finite universes of discourse provided by possibility measures Π_1 and Π_2 (with distributions π_1 and π_2), respectively. The possibility distribution π on $\mathbf{X}_1 \times \mathbf{X}_2$ is called *T-product possibility distribution* of π_1 and π_2 if for any $(x_1, x_2) \in \mathbf{X}_1 \times \mathbf{X}_2$

$$\pi(x_1, x_2) = T(\pi_1(x_1), \pi_2(x_2)). \tag{2}$$

Considering an arbitrary possibility distribution π defined on a product universe of discourse $\mathbf{X} \times \mathbf{Y}$, its *marginal possibility distribution* on $\mathbf{X}$ is defined by the equality

$$\pi_X(x) = \max_{y \in \mathbf{Y}} \pi(x, y) \tag{3}$$

for any $x \in \mathbf{X}$.

[1] max must be substituted by sup if $\mathbf{X}$ is not finite.

2.3 Conditioning

Let T be a t-norm on $[0,1]$. For any possibility measure Π on $\mathbf{X}$ with distribution π, we define the following binary relation on the set $\mathcal{G}(\mathbf{X}) = \{h : \mathbf{X} \longrightarrow [0,1]\}$ of all fuzzy variables on $\mathbf{X}$: For h_1 and h_2 in $\mathcal{G}(\mathbf{X})$ we say that h_1 and h_2 are (Π,T)-*equal almost everywhere* (and write $h_1 \overset{(\Pi,T)}{=} h_2$) if for any $x \in X$

$$T(h_1(x),\pi(x)) = T(h_2(x),\pi(x)).$$

This notion is very important for the definition of *conditional possibility distribution*, which is defined (in accordance with [2]) as *any* solution of the equation

$$\pi_{XY}(x,y) = T(\pi_Y(y), \pi_{X|_T Y}(x|_T y)), \tag{4}$$

for any $(x,y) \in \mathbf{X} \times \mathbf{Y}$. Continuity of a t-norm T guarantees the existence of a solution of this equation. This solution is not unique (in general), but the ambiguity vanishes when almost-everywhere equality is considered. We are able to obtain a representative of these conditional possibility distributions (if T is a continuous t-norm) by taking the residual $\pi_{XY}(x,\cdot)\triangle_T\pi_Y(\cdot)$ since

$$\pi_{X|_T Y}(x|_T \cdot) \overset{(\Pi_Y,T)}{=} \pi_{XY}(x,\cdot)\triangle_T\pi_Y(\cdot). \tag{5}$$

This way of conditioning brings a unifying view on several conditioning rules [4, 5, 7], i.e., its importance from the theoretical viewpoint is obvious. On the other hand, its practical meaning is not so substantial. Although De Cooman [2] claims that conditional distributions are never used *per se*, there exist situations in which it is necessary to be careful to choose an appropriate representative of the set of solutions (cf. Example 5 in [14]). Therefore, in this contribution we also use residuals rather than general conditionals.

2.4 Independence

Two variables X and Y (taking their values in $\mathbf{X}$ and $\mathbf{Y}$, respectively) are *possibilistically T-independent* [2] if for any $F_X \in X^{-1}(\mathcal{P}(\mathbf{X}))$, $F_Y \in Y^{-1}(\mathcal{P}(\mathbf{Y}))$,

$$\begin{aligned}
\Pi(F_X \cap F_Y) &= T(\Pi(F_X),\Pi(F_Y)),\\
\Pi(F_X \cap F_Y^C) &= T(\Pi(F_X),\Pi(F_Y^C)),\\
\Pi(F_X^C \cap F_Y) &= T(\Pi(F_X^C),\Pi(F_Y)),\\
\Pi(F_X^C \cap F_Y^C) &= T(\Pi(F_X^C),\Pi(F_Y^C)),
\end{aligned}$$

where A^C denotes the complement of A.

From this definition it immediately follows that the independence notion is parameterized by T. More specifically, it means that if X and Y are independent with respect to Gödel's t-norm, they need not be, for example, independent with respect to product t-norm. This fact is reflected in most definitions and assertions that follow.

In [11] we generalized this notion and in the following way: Given a possibility measure Π on $\mathbf{X} \times \mathbf{Y} \times \mathbf{Z}$ with the respective distribution $\pi(x,y,z)$, variables X and Y are *possibilistically conditionally T-independent*[2] given Z (in symbols $I_T(X,Y|Z)$) if, for any pair$(x,y) \in \mathbf{X} \times \mathbf{Y}$,

$$\pi_{XY|_T Z}(x,y|_T \cdot) \overset{(\Pi_Z,T)}{=} T(\pi_{X|_T Z}(x|_T \cdot), \pi_{Y|_T Z}(y|_T \cdot)). \tag{6}$$

Let us stress again that we do not deal with the pointwise equality but with the *almost everywhere equality*. This definition unifies, in a sense, several notions of conditional noninteractivity and that of conditional independence (for more details see [12]). Although it may seem to be controversial from the epistemic point of view [1], it is very suitable for our purpose, since it is closely connected (for more details see [13]) with a principal notion of multidimensional models — the notion of factorization.

We will say that a possibility distribution π *factorizes*[3] with respect to a system $\mathcal{A}$ and a t-norm T, if, for all complete subsets $A \in \mathcal{A}$, there exist fuzzy variables f_A of x_A such that π has the form

$$\pi(x) = T^{|\mathcal{A}|}(f_{A_1}(x_{A_1}), \ldots, f_{A_{|\mathcal{A}|}}(x_{A_{|\mathcal{A}|}})). \tag{7}$$

The functions f_A are not uniquely determined (in general), since they can be "multiplied" in several ways, cf. Example 14 in [13].

3 Iterative Proportional Fitting Procedure

In this section we define (in the most general way) an iterative proportional fitting procedure for possibility distributions and show, on a simple example, how it works.

Before doing that, let us recall what is possiblistic marginal problem.

3.1 Possibilistic Marginal Problem

Let us assume that $\mathbf{X}_i$, $i \in N$, $1 \leq |N| < \infty$ are finite universes of discourse, $\mathcal{K}$ is a system of nonempty subsets of N and $\mathcal{S} = \{\pi_K, K \in \mathcal{K}\}$ is a family of possibility distributions, where each π_K is a distribution on a product space

$$\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i.$$

[2] Let us note that a similar definition of conditional independence can be found in [8].

[3] Factorization is usually defined with respect to a graph, but this definition is more appropriate for the purpose of this contribution.

The problem we are interested in is the existence of an *extension*, i.e., a distribution π on

$$\mathbf{X} = \times_{i \in N} \mathbf{X}_i.$$

whose marginals are distributions from $\mathcal{S}$; or, more generally, the set

$$\mathcal{P} = \{\pi(x) : \pi(x_K) = \pi_K(x_K), K \in \mathcal{K}\}$$

is of interest.

The necessary condition (but not sufficient, as shown in [14]) for the existence of an extension is the pairwise projectivity of distributions from $\mathcal{S}$. Let us recall that two possibility distributions π_I and π_J are *projective* if they have common marginals, i.e. if

$$\pi_I(x_{I \cap J}) = \pi_J(x_{I \cap J}).$$

Since IPFP is able to solve a marginal problem (if a solution exists) within a probabilistic setting, it seems to be useful to design an analogous procedure for possibility distributions.

3.2 Design of Iterative Proportional Fitting Procedure

Let $\mathcal{S} = \{\pi_i, i = 1, \ldots m\}$ be a sequence of low-dimensional normal possibility distributions, which will be referred to as an *input sequence*. Let

$$\rho_{(0)} \in \mathcal{R} = \{\rho : \mathbf{X} \longrightarrow [0,1]; \max_{x \in \mathbf{X}} \rho(x) = 1\}$$

be an *initial possibility distribution*.

The *iterative proportional fitting procedure with respect to a t-norm T* (IPFP(T)) is a computational process defined for $x \in \mathbf{X}$ and for $j = 1, 2, \ldots$ and $k = (((j-1) \bmod m) + 1)$ by the following formula:

$$\rho_{(j)}(x) \quad = \quad T(\rho_{(j-1)}(x) \triangle_T \rho_{(j-1)}(x_{K_k}), \pi_k(x_{K_k})). \tag{8}$$

Formula (8) has the following meaning: at every step j we udate distribution $\rho_{(j-1)}$ simply by "multiplying" the marginal π_k, $k = (((j-1) \bmod m) + 1)$ by the residual of $\rho_{(j-1)}$ in order to obtain distribution $\rho_{(j)}$ such that

$$\rho_{(j)}(x_{K_k}) = \pi_k(x_{K_k}).$$

It is completely analogous to probability theory, where (8) has form

$$Q_{(j)}(x) = P_k(x_{K_k}) \frac{Q_{(j-1)}(x)}{Q_{(j-1)}(x_{K_k})},$$

which is a generalization of the original procedure by Deming and Stephan [3].

3.3 Example

The following simple example illustrates how the computations of IPFP(T) are performed.

Example 2 Let X_1, X_2 and X_3 be three binary variables with values in $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$, respectively ($\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3 = \{0,1\}$), and let the input sequence consist of two possibility distributions $\pi_1(x_1,x_2)$ and $\pi_2(x_2,x_3)$ on $\mathbf{X}_{\{1,2\}}$ and $\mathbf{X}_{\{2,3\}}$, respectively.

- The initial distribution $\rho_{(0)} \in \mathcal{R}$ is the least informative distribution on $\mathbf{X}_{\{1,2,3\}}$, i.e. $\rho_{(0)} \equiv 1$ (initial and input distributions can be found at Figure 1).

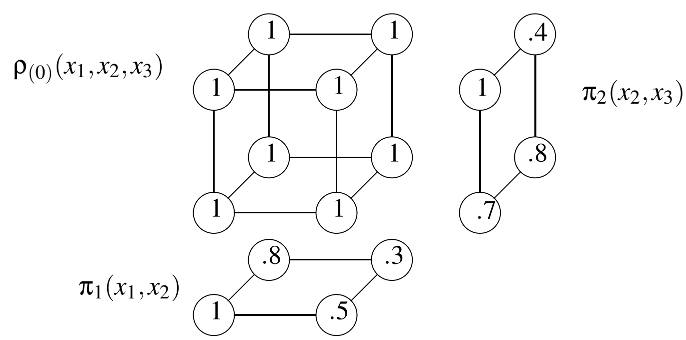

Figure 1: Initial and input distributions of IPFP(T)

- The operation of *fitting* the first input distribution $\pi_1(x_1,x_2)$ brings joint possibility distribution $\rho_{(1)}$ such that

$$\rho_{(1)}(x_1,x_2) = \pi_1(x_1,x_2),$$

as can be seen from Figure 2.

- *Fitting* the second input distribution $\pi_2(x_2,x_3)$ gives the joint possibility distribution

$$\rho_2(x_1,x_2,x_3) = T(\pi_2(x_2,x_3), \rho_{(1)}(x_1,x_2,x_3) \triangle_T \rho_{(1)}(x_2,x_3))$$

with the property $\rho_{(2)}(x_2,x_3) = \pi_2(x_2,x_3)$ (cf. Figure 3).

From Figure 3 one can see that due to the projectivity of π_1 and π_2, $\rho_{(2)}$ preserves its marginal from previous step, i.e.

$$\rho_{(2)}(x_1,x_2) = \rho_{(1)}(x_1,x_2) = \pi_1(x_1,x_2).$$

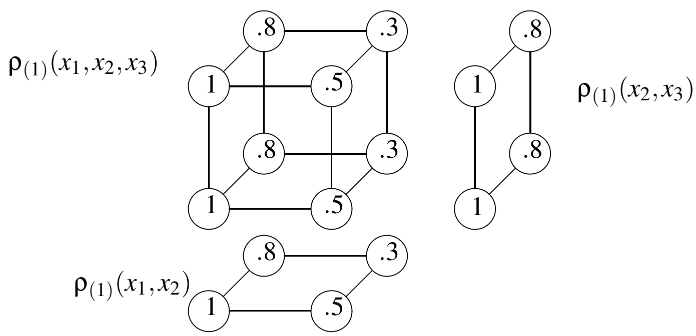

Figure 2: Joint distribution $\rho_{(1)}$ and its marginals after fitting π_1

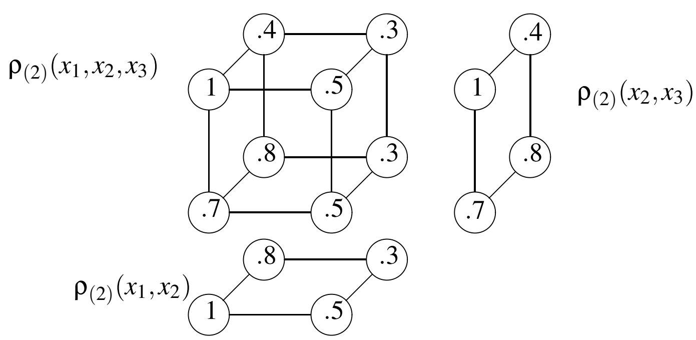

Figure 3: Joint distribution $\rho_{(2)}$ (with respect to Gödel's t-norm) and its marginals after fitting π_1 and π_2

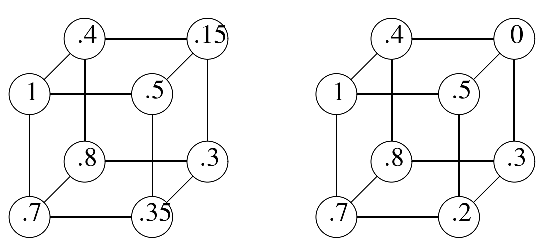

Figure 4: Joint distribution $\rho_{(2)}$ (with respect to product and Lukasziewicz' t-norms, respectively) after fitting π_1 and π_2

It is evident that there is no reason to fit $\pi_1(x_1, x_2)$ again, since it cannot bring any change to $\rho_{(2)}$.

From this simple example, one can conclude that if the input set consists of two projective possibility distributions and the initial possibility distribution is $\rho_{(0)} \equiv 1$, IPFP(T) stops after one cycle for any continuous t- norm T. Nevertheless, the resulting distribution depends on the choice of the t-norm, which can be seen from Figures 3 and 4.

4 On Convergence of Possibilistic IPFP

In this section we will generalize the observation from the end of the foregoing section and find a sufficient condition for the convergence of possibilistic IPFP. Before doing that, let us briefly recall the notions of operators of composition of possibility distributions (introduced in [10]), which seem to be a useful technical tool for proofs.

4.1 Operators of Composition

Considering a continuous t-norm T, two subsets K_1, K_2 of N and two normal possibility distributions $\pi_1(x_{K_1})$ and $\pi_2(x_{K_2})$,[4] we define the *operator of right composition* of these possibilistic distributions by the expression

$$\pi_1\left(x_{K_1}\right) \triangleright_T \pi_2\left(x_{K_2}\right) = T\left(\pi_1\left(x_{K_1}\right), \pi_2\left(x_{K_2}\right) \triangle_T \pi_2\left(x_{K_1 \cap K_2}\right)\right);$$

analogously the *operator of left composition* is defined by the expression

$$\pi_1\left(x_{K_1}\right) \triangleleft_T \pi_2\left(x_{K_2}\right) = T\left(\pi_1\left(x_{K_1}\right) \triangle_T \pi_1\left(x_{K_1 \cap K_2}\right), \pi_2\left(x_{K_2}\right)\right).$$

If $K_1 \cap K_2 = \emptyset$ then obviously

$$\pi_1\left(x_{K_1}\right) \triangleright_T \pi_2\left(x_{K_2}\right) = \pi_1\left(x_{K_1}\right) \triangleleft_T \pi_2\left(x_{K_2}\right) = T\left(\pi_1\left(x_{K_1}\right), \pi_2\left(x_{K_2}\right)\right),$$

which means that the operators of composition generalize, in a sense, T- product possibility distributions defined by (2).

It is evident that both $\pi_1 \triangleright_T \pi_2$ and $\pi_1 \triangleleft_T \pi_2$ are (generally different) possibility distributions of variables $(X_i)_{i \in K_1 \cup K_2}$. In fact, the first one is an extension of π_1, while the second of π_2, in a special case of both, as the following lemma suggests.

Lemma 1 *Consider two distributions $\pi_1(x_{K_1})$ and $\pi_2(x_{K_2})$. Then*

$$(\pi_1 \triangleright_T \pi_2)(x_{K_1 \cup K_2}) = (\pi_1 \triangleleft_T \pi_2)(x_{K_1 \cup K_2})$$

for any continuous t-norm T if and only if π_1 and π_2 are projective.

[4] Let us stress that for the definition of these operators we do not require projectivity of distributions π_1 and π_2.

The following lemma (proven in [14]) expresses the relationship between the operators of composition and conditional T-independence.

Lemma 2 *Let T be a continuous t-norm and π_1 and π_2 be projective possibility distributions on $\mathbf{X}_{K_1}$ and $\mathbf{X}_{K_2}$, respectively. Then the distribution π of $X_{K_1 \cup K_2}$*

$$\pi(x_{K_1 \cup K_2}) = \pi_1(x_{K_1}) \triangleright_T \pi_2(x_{K_2}) = \pi_1(x_{K_1}) \triangleleft_T \pi_2(x_{K_2})$$

if and only if $X_{K_1 \setminus K_2}$ and $X_{K_2 \setminus K_1}$ are conditionally independent, given $X_{K_1 \cap K_2}$.

4.2 Perfect Sequences

Now, we will recall how to apply the operators iteratively. Consider a sequence of distributions $\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_m(x_{K_m})$ and the expression

$$\pi_1 \triangleright_T \pi_2 \triangleright_T \ldots \triangleright_T \pi_m.$$

Before presenting its properties, let us note that in the part that follows, we always apply the operators from left to right, i.e.,

$$\pi_1 \triangleright_T \pi_2 \triangleright_T \pi_3 \triangleright_T \ldots \triangleright_T \pi_m = (\ldots((\pi_1 \triangleright_T \pi_2) \triangleright_T \pi_3) \triangleright_T \ldots \triangleright_T \pi_m).$$

This expression defines a multidimensional distribution on $\mathbf{X}_{K_1 \cup \ldots \cup K_m}$. Therefore, for any permutation $i_1, i_2, \ldots, i_m$ of indices $1, \ldots, m$ the expression

$$\pi_{i_1} \triangleright_T \pi_{i_2} \triangleright_T \ldots \triangleright_T \pi_{i_m}$$

determines a distribution on the same universe of discourse. However, for different permutations these distributions can differ from one another. Some of them seem to possess the most advantageous properties.

An ordered sequence of possibility distributions $\pi_1, \pi_2, \ldots, \pi_m$ is said to be *T-perfect* if for any $j = 2, \ldots, m$

$$\pi_1 \triangleright_T \cdots \triangleright_T \pi_j = \pi_1 \triangleleft_T \cdots \triangleleft_T \pi_j.$$

The notion of T-perfectness suggests that a sequence perfect with respect to one t-norm needn't be perfect with respect to another t-norm, similarly to (conditional) T-independence.

Let us present two assertions, which will be used later.

Lemma 3 *Let T be a continuous t-norm. The sequence $\pi_1, \pi_2, \ldots, \pi_m$ is T-perfect, if and only if the pairs of distributions $(\pi_1 \triangleright_T \cdots \triangleright_T \pi_{k-1})$ and π_k are projective for all $k = 2, 3, \ldots, m$.*

Theorem 1 *The sequence $\pi_1, \pi_2, \ldots, \pi_m$ is T-perfect if and only if all the distributions $\pi_1, \pi_2, \ldots, \pi_m$ are marginal to distribution $\pi_1 \triangleright_T \pi_2 \triangleright_T \ldots \triangleright \pi_m$.*

Now, let us recall the notion of *running intersection property* (RIP) and the related results from [14]. A sequence of sets $K_1, K_2, \ldots, K_n$ is said to meet RIP if

$$\forall i = 2, \ldots, n \;\; \exists j (1 \le j < i) \qquad (K_i \cap (K_1 \cup \ldots \cup K_{i-1})) \subseteq K_j.$$

Lemma 4 *If $\pi_1, \pi_2, \ldots, \pi_m$ is a sequence of pairwise projective low-dimensional distributions such that $K_1, \ldots, K_m$ meets RIP, then this sequence is T-perfect for any continuous t-norm T.*

4.3 Convergence of IPFP(T)

Theorem 2 *If there is an ordering $\pi_1, \ldots, \pi_m$ of possibility distributions from $\mathcal{S}$ such that $\pi_1, \ldots, \pi_m$ form a T-perfect sequence for some continuous t-norm T and $\rho_{(0)} \equiv 1$, then IPFP(T) converges in one cycle. Furthermore, distribution $\rho_{(m)}$ factorizes with respect to $\mathcal{K}$ and T.*

Proof. First, let us note that (8) for $\pi_1, \ldots, \pi_m$ can be rewritten using an operator of left composition, i.e.,

$$\begin{aligned} \rho_{(j)}(x) &= T(\rho_{(j-1)}(x) \triangle_T \rho_{(j-1)}(x_{K_k}), \pi_k(x_{K_k})) \\ &= \rho_{(j-1)}(x) \triangleleft_T \pi_k(x_{K_k}) \end{aligned}$$

for any $j = 1, \ldots$; especially for $j = 1, \ldots, n$ (which means that $k = j$) we obtain

$$\begin{aligned} \rho_{(j)}(x) &= \rho_{(j-1)}(x) \triangleleft_T \pi_j(x_{K_j}) \\ &= (\rho_{(j-2)}(x) \triangleleft_T \pi_{j-1}(x_{K_{j-1}})) \triangleleft_T \pi_j(x_{K_j}) \\ &\ldots \\ &= (\ldots(\rho_{(0)}(x) \triangleleft_T \pi_1(x_{K_1})) \triangleleft_T \ldots \triangleleft_T \pi_{j-1}(x_{K_{j-1}})) \triangleleft_T \pi_j(x_{K_j}) \\ &= (\ldots T(\rho_{(0)}(x_{N \setminus \cup_{k=1}^{j} K_k}), \pi_1(x_{K_1})) \triangleleft_T \ldots \triangleleft_T \pi_{j-1}(x_{K_{j-1}})) \triangleleft_T \pi_j(x_{K_j}), \end{aligned}$$

since $\rho_{(0)} \equiv 1$. In particular we have

$$\rho_{(m)}(x) = (\ldots(\pi_1(x_{K_1}) \triangleleft_T \pi_2(x_{K_2}) \ldots \triangleleft_T \pi_{m-1}(x_{K_{m-1}})) \triangleleft_T \pi_m(x_{K_m}). \tag{9}$$

Since $\pi_1, \ldots, \pi_m$ is a T-perfect sequence of possibility distributions, every π_k is a marginal to the distribution on the right-hand side of (9). Therefore,

$$\rho_{(m)}(x_{K_k}) = \pi_k(x_{K_k})$$

for all $k = 1, \ldots, m$, which implies

$$\rho_{(j)}(x_{K_k}) = \rho_{(m)}(x_{K_k})$$

for any $j = m+1, \ldots$. To prove factorization it is enough to find fuzzy variables $f_{K_1}, \ldots, f_{K_m}$ such that

$$\rho_{(m)}(x) = T^m(f_{K_1}(x_{K_1}), \ldots, f_{K_m}(x_{K_m})).$$

But, due to T-perfectness of $\pi_1, \ldots, \pi_m$

$$\rho_{(m)}(x) = \pi_1(x_{K_1}) \triangleright_T \pi_2(x_{K_2}) \triangleright_T \ldots \triangleright \pi_m(x_{K_m}),$$

which can be rewritten in the form

$$\rho_{(m)}(x) = T^m(\pi_1(x_{K_1}), \pi_2(x_{K_2}) \triangle_T \pi_2(x_{K_2 \cap K_1}), \ldots \\ \ldots, \pi_m(x_{K_m}) \triangle_T \pi_m(x_{K_m \cap (K_1 \cup \ldots \cup K_{m-1})})),$$

which concludes the proof.

First, let us stress that perfectness with respect to a t-norm implies convergence with respect to the same t-norm (and *not with respect to any*) as can be seen from the following simple example.

Example 3 Let X_1, X_2 and X_3 be three binary variable as in Example 2 and π_1, π_2 and π_3 on $\mathbf{X}_{\{1,2\}}, \mathbf{X}_{\{2,3\}}$ and $\mathbf{X}_{\{1,3\}}$ be defined by Table 1.

π_1 X_2	0	1
$X_1 = 0$	1	.8
$X_1 = 1$	.6	.4

π_2 X_3	0	1
$X_2 = 0$	1	.5
$X_2 = 1$	.3	.8

π_3 X_3	0	1
$X_1 = 0$	1	.8
$X_1 = 1$	.6	.5

Table 1: Distributions forming min-perfect sequence

$\rho_{(j)}$	j						
(x_1, x_2, x_3)	0	1	2	3	4	5	6
$(0,0,0)$	1	1	1	1	1	1	1
$(0,0,1)$	1	1	.5	.5	.5	.5	.5
$(0,1,0)$	1	.8	.8	.8	.8	.8	.8
$(0,1,1)$	1	.8	.8	.8	.8	.8	.8
$(1,0,0)$	1	.6	.6	.6	.6	.6	.6
$(1,0,1)$	1	.6	.5	.5	.5	.5	.5
$(1,1,0)$	1	.4	.3	.3	.3	.3	.3
$(1,1,1)$	1	.4	.4	.4	.4	.4	.4

Table 2: Convergence of IPFP with respect to Gödel's t-norm

Sequence π_1, π_2, π_3 is min-perfect (due to Lemma 3), since $\pi_1(x_2) = \pi_2(x_2)$ and $(\pi_1 \triangleright_{T_G} \pi_2)(x_1, x_3) = \pi_3(x_1, x_3)$. Starting from $\rho_{(0)} \equiv 1$, IPFP(T_G) converges after one cycle as can be seen from Table 2 while IPFP(T_Π) and IPFP(T_L) converge after four and five cycles, respectively (cf. Tables 3 and 4).

Corollary 1 *If there is a permutation $K_{i_1}, \ldots, K_{i_n}$ of sets from $\mathcal{K}$ such that $K_{i_1}, \ldots, K_{i_n}$ meets RIP, $\{\pi_{i_1}, \ldots, \pi_{i_n}\}$ is an input sequence of pairwise projective possibility distributions and $\rho_{(0)} \equiv 1$ then IPFP(T) converges in one cycle for any continuous t-norm T and $\rho_{(m)}$ factorizes with respect to the corresponding t-norm T.*

$\rho_{(j)}$	j								
(x_1,x_2,x_3)	0	1	2	3	4, 5	6	7, 8	9	10, 11, 12
$(0,0,0)$	1	1	1	1	1	1	1	1	1
$(0,0,1)$	1	1	.5	.5	.5	.5	.5	.5	.5
$(0,1,0)$	1	.8	.3	.3	.3	.3	.3	.3	.3
$(0,1,1)$	1	.8	.8	.8	.8	.8	.8	.8	.8
$(1,0,0)$	1	.6	.6	.6	.6	.6	.6	.6	.6
$(1,0,1)$	1	.6	.3	.375	.375	.46875	.46875	.5	.5
$(1,1,0)$	1	.4	.15	.15	.12	.096	.096	.096	.09
$(1,1,1)$	1	.4	.4	.5	.4	.5	.4	.427	.4

Table 3: Convergence of IPFP with respect to product t-norm

$\rho_{(j)}$	j										
(x_1,x_2,x_3)	0	1	2	3	4, 5	6	7, 8	9	10, 11	12	13, 14, 15
(0,0,0)	1	1	1	1	1	1	1	1	1	1	1
(0,0,1)	1	1	.5	.5	.5	.5	.5	.5	.5	.5	.5
(0,1,0)	1	.8	.3	.3	.3	.3	.3	.3	.3	.3	.3
(0,1,1)	1	.8	.8	.8	.8	.8	.8	.8	.8	.8	.8
(1,0,0)	1	.6	.6	.6	.6	.6	.6	.6	.6	.6	.6
(1,0,1)	1	.6	.1	.2	.2	.3	.3	.4	.4	.5	.5
(1,1,0)	1	.4	.0	.0	.0	.0	.0	.0	.0	.0	.0
(1,1,1)	1	.4	.4	.5	.4	.5	.4	.5	.4	.5	.4

Table 4: Convergence of IPFP with respect to Lukasziewicz' t-norm

Proof follows directly from Theorem 2 and Lemma 4.

Let us also mention that $\rho_{(0)} \equiv 1$ is not only a technical requirement that makes the proof of Theorem 2 so simple; it may be substantial for convergence as can be seen from the following example.

Example 4 Let $\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3$ and $\pi_1(x_1,x_2)$, $\pi_2(x_2,x_3)$ be as in Example 2 and $\rho_{(0)}$ be defined as follows:

$$\rho_{(0)}(0,0,0) = \rho_{(0)}(0,1,1) = \rho_{(0)}(1,0,1) = \rho_{(0)}(1,1,0) = 1,$$

values of remaining combinations being equal to $\alpha \in [0,1]$. The convergence depends on the value of α — the results of our experiments can be found in Table 5.

The reason for this behaviour lies in the tendency of IPF procedure to find a distribution with given marginals which, moreover, factorizes with respect to the system $\mathcal{K}$ and is "as close as possible" to $\rho_{(0)}$. It is evident that $\rho_{(0)} \equiv 1$ factorizes with respect to *any* system of cliques. Therefore, it is the "safe", although perhaps

α	Convergence of IPFP(T)		
	T_G	T_Π	T_L
1	1	1	1
.5	2	2	2
.1	cycles	4	3
0	cycles	cycles	3

Table 5: Convergence of IPFP(T) depends on α

not always an optimal, initial distribution. The more “distant” the structure of the starting distribution is from factorization with respect to $\mathcal{K}$ and T, the more problematic the convergence of IPFP(T) is.

5 Conclusions

We introduced a possibilistic version of IPF procedure with the aim of using it as a tool for marginal problem solving. This procedure is parameterized by a continuous t-norm and its behaviour (convergence) is strongly dependent on it. Another important finding is that convergence of IPFP(T) substantially depends on the choice of an input distribution.

Nevertheless, there are still many problems that remain to be solved. The most important is the proof of the convergence of IPFP(T) in a general case. Another question is whether the resulting distribution is independent of the ordering of input distributions. We should also study the behaviour of IPFP(T) in inconsistent cases.

Acknowledgements

The author wishes to express her gratitude to Jiří Vomlel for writing the program in Mathematica and for his invaluable help when performing the experiments.

References

[1] B. Bouchon-Meunier, G. Coletti and C. Marsala, Independence and possibilistic conditioning *Annals of Mathematics and Artificial Intelligence,* **35** (2002), pp. 107-123.

[2] G. de Cooman, Possibility theory I – III. *Int. J. General Systems* **25** (1997), pp. 291–371.

[3] W.E. Deming and F.F. Stephan, On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11** (1940), pp. 427–444.

[4] A. Dempster, Upper and lower probabilities induced by multivalued mappings. *Ann. Math. Statist.* **38** (1967), pp. 325–339.

[5] D. Dubois and H. Prade, *Possibility theory.* Plenum Press, New York, 1988.

[6] P. Hájek, T. Havránek and R. Jiroušek, *Uncertain information processing in expert systems.* CRC Press, Boca Raton, Ann Arbor, London, Tokyo, 1992.

[7] E. Hisdal, Conditional possibilities independence and noninteraction, *Fuzzy Sets and Systems*, **1** (1978), pp. 299–309.

[8] H. Janssen, G. de Cooman and E. E. Kerre, First results for a mathematical theory of possibilistic Markov processes, *Proceedings of IPMU'96, volume III (Information Processing and Management of Uncertainty in Knowledge-Based Systems),* Granada, Spain, 1996, pp. 1425–1431. (1988),

[9] L. Rüschendorf, Convergence of the iterative proportional fitting procedure. *Ann. Statist.* **24** (1995), pp. 1160–1174.

[10] J. Vejnarová, Composition of possibility measures on finite spaces: preliminary results. In: *Proceedings of IPMU'98,* Paris, 1998, pp. 25–30.

[11] J. Vejnarová, Possibilistic independence and operators of composition of possibility measures. In: M. Hušková, J. Á. Víšek, P. Lachout (eds.) *Prague Stochastics'98,* JČMF, 1998, pp. 575–580.

[12] J. Vejnarová, Conditional independence relations in possibility theory. *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **8** (2000), pp. 253–269.

[13] J. Vejnarová, Markov properties and factorization of possibility distributions. *Annals of Mathematics and Artificial Intelligence,* **35** (2002), pp. 357-377.

[14] J. Vejnarová, A partial solution of possibilistic marginal problem. In: G. de Cooman, T. L. Fine and T. Seidenfeld (eds.), *ISIPTA'01 (2nd Int. Symposium on Imprecise Probabilities and Their Applications)*, Shaker Publishing BV, Maastricht, 2001, pp. 342–351.

[15] L. A. Zadeh, Fuzzy sets as a basis for theory of possibility, *Fuzzy Sets and Systems,* **1** (1978), pp. 3–28.

Jiřina Vejnarová is with the Laboratory for Intelligent Systems of the University of Economics, Prague, Czech Republic. She is also a senior research fellow of the Institute for Information Theory and Automation of the Academy of Sciences of the Czech Republic. E-mail: vejnar@vse.cz