Kybernetika

Radim Jiroušek

A short note on Perez's approximation by dependence structure simplification

Kybernetika, Vol. 43 (2007), No. 5, 649--656

Persistent URL: http://dml.cz/dmlcz/135804

KYBERNETIKA — VOLUME *43* (2007), NUMBER 5, PAGES 649 – 656

A SHORT NOTE ON PEREZ'S APPROXIMATION BY DEPENDENCE STRUCTURE SIMPLIFICATION

Radim Jiroušek

Perez's approximations of probability distributions by dependence structure simplification were introduced in 1970s, much earlier than graphical Markov models. In this paper we will recall these Perez's models, formalize the notion of a compatible system of elementary simplifications and show the necessary and sufficient conditions a system must fulfill to be compatible. For this we will utilize the apparatus of compositional models.

Keywords: approximation of probability distributions, dependence structure simplification, compatibility, compositional models.

AMS Subject Classification: 62E17, 68T30

1. INTRODUCTION

As early as in 1960s Albert Perez conceived that when considering probabilistic models of practical problems one has to simplify them, to *approximate* them, otherwise one easily gets beyond the boundary of computational tractability. Therefore he started to study problems of *data reduction* [3] (see also paper [6] of this special issue) and problems of ε-sufficiency of probability distributions. In this context he published in 1977 his perhaps the most cited paper: *ε-admissible simplifications of dependence structure of a set of random variables* [5]. In this paper he used the notion of *dependence structure simplification* approximation for the first time.

Later in 1980s, when he started being interested in probabilistic tools for expert systems, Perez denoted by the term *dependence structure simplification* the class of models which could easily be computed from a system of oligodimensional distributions representing pieces of partial knowledge. In this second application, the problems of compatibility of global and local knowledge played an important role and highlighted thus the importance to find necessary and sufficient conditions guaranteeing the compatibility. For a solution of this problem we shall use the notation usual in the field of *compositional models*, whose origination was also inspired by Perez's ideas.

2. ELEMENTARY SIMPLIFICATION

Consider a system of finite-valued random variables X_i with indices from a non-empty finite set N. All the probability distributions discussed in the paper will be denoted by Greek letters. For $K \subset N$, $\kappa(K)$ denotes a distribution of variables

$$X_K = \{X_i\}_{i \in K},$$

and $\kappa(x_K)$ denotes its value for the vector x_K from a Cartesian product $\times_{i \in K} \boldsymbol{X}_i$ ($\boldsymbol{X}_i$ is the set of values of variable X_i). Having $L \subset K$ and a distribution $\kappa(K)$, we will denote its corresponding marginal distribution either $\kappa(L)$, or $\kappa^{\downarrow L}$. For the same L, K and $x \in \times_{i \in K} \boldsymbol{X}_i$, x_L denotes the projection of x into $\times_{i \in L} \boldsymbol{X}_i$.

Compositional models were first introduced in [1]. They are based on application of a simple operator of composition creating from two probability distributions a new one, which is defined for the variables appearing among the arguments of at least one from the original distributions:

Definition 1. For two arbitrary distributions $\kappa(K)$ and $\lambda(L)$ their *composition* is given by the formula

$$\kappa \triangleright \lambda = \begin{cases} \frac{\kappa \cdot \lambda}{\lambda^{\downarrow K \cap L}} & \text{when} \quad \kappa^{\downarrow K \cap L} \ll \lambda^{\downarrow K \cap L}, \\ \text{undefined} & \text{otherwise}, \end{cases}$$

where the symbol $\kappa^{\downarrow M} \ll \lambda^{\downarrow M}$ denotes that $\kappa^{\downarrow M}$ is *dominated* by $\lambda^{\downarrow M}$, which means (in the considered finite setting)

$$\forall x_M \in \times_{i \in M} \boldsymbol{X}_i \quad (\lambda(x_M) = 0 \implies \kappa(x_M) = 0).$$

Remark. If the marginal $\lambda^{\downarrow K \cap L}$ dominates $\kappa^{\downarrow K \cap L}$ then the formula in the definition is evaluated point-wise, i. e., for each $x \in \boldsymbol{X}_{K \cup L}$ value

$$(\kappa \triangleright \lambda)(x) = \frac{\kappa(x_K) \cdot \lambda(x_L)}{\lambda(x_{K \cap L})}$$

is computed.

Using this operator, Perez's *elementary* (E, D)*-simplification*[1] of a distribution $\kappa(K)$ (for $D \subset E \subset K$) is a distribution

$$\bar{\kappa} = \kappa^{\downarrow E} \triangleright \kappa^{\downarrow K \setminus D}.$$

In connection with this formula it is important to realize that because the operator $\triangleright$ is applied to two marginal distributions of κ, elementary (E, D)-simplification is always defined. Moreover, since $E \cup (K \setminus D) = K$, distribution $\bar{\kappa}$ is defined for the same set of variables as κ.

For the reader not familiar with the basic properties of the operator of composition we include two short passages recollecting the most important properties that were proved already in the papers [1, 2].

[1] When comparing this text with the original paper [5] notice that Perez speaks about elementary (E, F)-simplification, where $F = E \setminus D$. We adopted this small change of notation in order to simplify some of the formulas.

Basic properties of the operator of composition I

Lemma 1. Consider two distributions $\kappa(K)$ and $\lambda(L)$ for which the composition $\kappa \triangleright \lambda$ is defined. Then

1. $(\kappa \triangleright \lambda)^{\downarrow K} = \kappa$.
2. $\kappa \triangleright \lambda = \lambda \triangleright \kappa \iff \kappa^{\downarrow K\cap L} = \lambda^{\downarrow K\cap L}$.
3. For the distribution $\kappa \triangleright \lambda$, groups of variables $X_{K\setminus L}$ and $X_{L\setminus K}$ are conditionally independent given variables $X_{K\cap L}$. This will be in the following text expressed by the symbol
$$X_{K\setminus L} \perp\!\!\!\perp X_{L\setminus K} | X_{K\cap L} \; [\kappa \triangleright \lambda].$$

Remark. From Property 2 of this Lemma we can see that the operator of composition is not commutative. It is not difficult to see that this operator is neither associative[2]. Therefore if we consider multiple applications of the operator we have to specify in which order they should be performed. To make the formulas more lucid we will omit brackets in case that the operator is to be applied from left to right, i. e., in what follows

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 \triangleright \ldots \triangleright \kappa_{n-1} \triangleright \kappa_n = (\ldots((\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3) \triangleright \ldots \triangleright \kappa_{n-1}) \triangleright \kappa_n.$$

Consider distribution $\bar{\kappa}$, which is an (E, D)-simplification of κ. From Properties 1 and 2. of Lemma 1 we immediately see that if either $D = \emptyset$ or $E = K$ then $\bar{\kappa}$ equals κ. On the other hand if $\emptyset \neq D \subset E \subsetneq K$ then (due to Property 3 of Lemma 1)

$$\bar{\kappa} = \kappa \iff X_D \perp\!\!\!\perp X_{K\setminus E} | X_{E\setminus D} \; [\kappa].$$

3. SIMPLIFICATION OF A DEPENDENCE STRUCTURE

Essentially, in [5] Perez introduced a simplification of a dependence structure as a *cumulation of a certain number of compatible elementary simplifications*.

To express this idea more exactly, consider sequence of elementary simplifications

$$(E_0, D_0), (E_1, D_1), \ldots, (E_n, D_n)$$

such that

$$D_n \subset E_n \subsetneq E_{n-1}, D_{n-1} \subset E_{n-1} \subsetneq E_{n-2}, \ldots, D_1 \subset E_1 \subsetneq E_0, D_0 \subset E_0 \subsetneq K.$$

By $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-*simplification* of a dependence structure of the distribution $\kappa(K)$ Perez understood the distribution

$$\bar{\kappa}^n = \kappa^{\downarrow E_n} \triangleright \kappa^{\downarrow E_{n-1}\setminus D_n} \triangleright \kappa^{\downarrow E_{n-2}\setminus D_{n-1}} \triangleright \ldots \triangleright \kappa^{\downarrow E_1\setminus D_2} \triangleright \kappa^{\downarrow E_0\setminus D_1} \triangleright \kappa^{\downarrow K\setminus D_0}.$$

[2] The reader can easily show it by the example:

$$\Big(\kappa(\{1\}) \triangleright \lambda(\{2\})\Big) \triangleright \mu(\{1,2\}) \neq \kappa(\{1\}) \triangleright \Big(\lambda(\{2\}) \triangleright \mu(\{1,2\})\Big).$$

Namely, the equality in this expression holds true only in case that $\mu(\{1,2\}) = \mu(\{1\})\mu(\{2\})$.

It is a compositional model, i. e., multidimensional distribution obtained by an iterative application of the operator of composition to the so-called *generating sequence*

$$\kappa^{\downarrow E_n}, \kappa^{\downarrow E_{n-1}\setminus D_n}, \kappa^{\downarrow E_{n-2}\setminus D_{n-1}}, \ldots, \kappa^{\downarrow E_1\setminus D_2}, \kappa^{\downarrow E_0\setminus D_1}, \kappa^{\downarrow K\setminus D_0}.$$

To answer the question what are the properties of the resulting distribution $\bar{\kappa}^n$ we have again to recall some of the notions and results achieved in the field of compositional models.

Basic properties of the operator of composition II

The reader familiar with some papers on compositional models knows that one of the most important notions of this theory is that of a so-called *perfect sequence*.

Definition 2. *A generating sequence of probability distributions* $\kappa_1, \kappa_2, \ldots, \kappa_n$ *is called* perfect *if* $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ *is defined and*

$$\begin{aligned} \kappa_1 \triangleright \kappa_2 &= \kappa_2 \triangleright \kappa_1, \\ \kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 &= \kappa_3 \triangleright (\kappa_1 \triangleright \kappa_2), \\ &\vdots \\ \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n &= \kappa_n \triangleright (\kappa_1 \triangleright \ldots \triangleright \kappa_{n-1}). \end{aligned}$$

From this definition one can hardly see what are the properties of the perfect sequences; the main one is expressed by the following characterization theorem, which was proved in [2].

Theorem 1. A sequence of distributions $\kappa_1, \kappa_2, \ldots, \kappa_n$ is perfect *iff* all the distributions from this sequence are marginals of the distribution $\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n$.

From the practical point of view it is also important to have a tool enabling us to recognize whether a generating sequence is perfect or not. For this one can take advantage of the following assertion (for its proof see [1]).

Lemma 2. A sequence $\kappa_1(K_1), \kappa_2(K_2), \ldots, \kappa_n(K_n)$ is perfect *iff* the pairs of distributions κ_m and $(\kappa_1 \triangleright \ldots \triangleright \kappa_{m-1})$ are consistent, i. e. if

$$\kappa_m^{\downarrow K_m\cap(K_1\cup\ldots\cup K_{m-1})} = (\kappa_1 \triangleright \ldots \triangleright \kappa_{m-1})^{\downarrow K_m\cap(K_1\cup\ldots\cup K_{m-1})},$$

for all $m = 2, 3, \ldots, n$.

At the end of this paper we shall also need the following (almost trivial) assertion.

Theorem 2. Let a sequence of pairwise consistent distributions $\kappa_1(K_1), \ldots, \kappa_n(K_n)$ be such that $K_1, K_2, \ldots, K_n$ meets the well-known running intersection property:

$$\forall\, i = 2, 3, \ldots, n \;\; \exists\, j (1 \leq j < i) \quad \text{such that} \quad K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \subseteq K_j.$$

Then $\kappa_1, \kappa_2, \ldots, \kappa_n$ is perfect.

4. BASIC PROPERTIES OF SIMPLIFICATION

In [5], Perez introduced a *dependence tightness* of a distribution and in Theorem 1.2 (presented below as Theorem 3) expressed the loss of this value when substituting a distribution by its dependence structure simplification.

The *dependence tightness* of a distribution (later called by other authors also informational content, or multiinformation) is a relative entropy (crossentropy) of a distribution with respect to the product of its one-dimensional marginals:

$$I(\kappa(K)) = \sum_{x \in \times_{i \in K} \boldsymbol{X}_i} \kappa(x) \frac{\log \kappa(x)}{\prod\limits_{i \in K} \kappa(x_i)}.$$

Theorem 3. Consider $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure of distribution $\kappa(K)$. The loss of dependence tightness caused by this simplification is given by

$$\begin{aligned} I(\kappa) - I(\bar{\kappa}^n) = MI_\kappa(E_0; K \setminus E_0) &- MI_\kappa(E_0 \setminus D_0; K \setminus E_0) \\ &+ MI_\kappa(E_1; E_0 \setminus E_1) - MI_\kappa(E_1 \setminus D_1; E_0 \setminus E_1) + \ldots \\ &+ MI_\kappa(E_n; E_{n-1} \setminus E_n) - MI_\kappa(E_n \setminus D_n; E_{n-1} \setminus E_n), \end{aligned}$$

where $MI_\kappa(B; C)$ (for B, C disjoint) is a Shannon mutual information defined by

$$MI_\kappa(B; C) = \sum_{x \in \times_{i \in B \cup C} \boldsymbol{X}_i} \kappa(x) \frac{\log \kappa(x)}{\kappa(x_B)\kappa(x_C)}.$$

The validity of this Perez's Theorem is based on the important property that for $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure $\bar{\kappa}^n$ of distribution κ

$$\bar{\kappa}^n(B) = \kappa(B)$$

holds true for all $B = K \setminus D_0, E_0 \setminus D_1, E_1, \setminus D_2, \ldots, E_{n-1} \setminus D_n, E_n$. Due to Theorem 1 this can be expressed also in other words: it is necessary that generating sequence

$$\kappa^{\downarrow E_n}, \kappa^{\downarrow E_{n-1} \setminus D_n}, \kappa^{\downarrow E_{n-2} \setminus D_{n-1}}, \ldots, \kappa^{\downarrow E_1 \setminus D_2}, \kappa^{\downarrow E_0 \setminus D_1}, \kappa^{\downarrow K \setminus D_0}.$$

is perfect. Let us show by the following simple example that generally this sequence need not be perfect. We will show that $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure need not retain even one-dimensional marginal distribution of κ.

Example.

Consider 3-dimensional distribution κ from Table 1 and its $(\{1,2\},\emptyset;\{1\},\{1\})$-simplification of a dependence structure

$$\bar{\kappa}^1(\{1,2,3\}) = \kappa^{\downarrow\{1\}} \triangleright \kappa^{\downarrow\{2\}} \triangleright \kappa,$$

which is also contained in Table 1. (When showing that the distribution from the last column of Table 1 is really $\kappa^{\downarrow\{1\}} \triangleright \kappa^{\downarrow\{2\}} \triangleright \kappa$, notice that both $\kappa^{\downarrow\{1\}}$ and $\kappa^{\downarrow\{2\}}$ are uniform and therefore also $\kappa^{\downarrow\{1\}} \triangleright \kappa^{\downarrow\{2\}}$ is a uniform distribution of the respective variables.)

Table 1. Probability distributions.

X_1	X_2	X_3	κ	$\kappa^{\downarrow\{1\}} \triangleright \kappa^{\downarrow\{2\}} \triangleright \kappa$
0	0	0	3/32	1/16
0	0	1	9/32	3/16
0	1	0	3/32	3/16
0	1	1	1/32	1/16
1	0	0	3/32	3/16
1	0	1	1/32	1/16
1	1	0	3/32	1/16
1	1	1	9/32	3/16

From this Table we immediately see that $\kappa(\{3\}) = [3/8; 5/8]$, whereas $\bar{\kappa}^1(\{3\}) = = [1/2; 1/2]$.

Let us go back to the Perez's introduction of the simplification of the dependence structure as *a cumulation of a certain number of* **compatible** *elementary simplifications,* because now we are able to give an exact meaning to the notion of a sequence of compatible elementary simplifications.

Definition 3. $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure of distribution $\kappa(K)$ is *compatible* if either

1. $n = 0$, or

2. $(E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure of distribution $\kappa^{\downarrow E_0}$ is compatible and

$$\left(\kappa^{\downarrow E_n} \triangleright \kappa^{\downarrow E_{n-1}\setminus D_n} \triangleright \kappa^{\downarrow E_{n-2}\setminus D_{n-1}} \triangleright \ldots \triangleright \kappa^{\downarrow E_1\setminus D_2} \triangleright \kappa^{\downarrow E_0\setminus D_1}\right)^{\downarrow E_0\setminus D_0} = \kappa^{\downarrow E_0\setminus D_0}.$$

Theorem 4. $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure of distribution $\kappa(K)$ is compatible *iff* the generating sequence

$$\kappa^{\downarrow E_n}, \kappa^{\downarrow E_{n-1}\setminus D_n}, \kappa^{\downarrow E_{n-2}\setminus D_{n-1}}, \ldots, \kappa^{\downarrow E_1\setminus D_2}, \kappa^{\downarrow E_0\setminus D_1}, \kappa^{\downarrow K\setminus D_0} \tag{1}$$

is perfect.

Proof. For $n = 0$ any simplification is compatible so, we have just to show that $\kappa^{\downarrow E_0}, \kappa^{\downarrow K\setminus D_0}$ is perfect, but it directly follows from Property 2 of Lemma 1.

Consider $n \geq 1$ and assume the assertion holds true for $n-1$. If (1) is perfect then, due to Lemma 2

$$\left(\kappa^{\downarrow E_n} \triangleright \kappa^{\downarrow E_{n-1}\setminus D_n} \triangleright \kappa^{\downarrow E_{n-2}\setminus D_{n-1}} \triangleright \ldots \triangleright \kappa^{\downarrow E_1\setminus D_2} \triangleright \kappa^{\downarrow E_0\setminus D_1}\right)^{\downarrow (K\setminus D_0)\cap E_0} = \kappa^{\downarrow (K\setminus D_0)\cap E_0}.$$

$(E_1, D_1; \ldots; E_n, D_n)$-simplification of a dependence structure of distribution $\kappa(K)$ is compatible due to the inductive assumption, and therefore, since $(K \setminus D_0) \cap E_0 = E_0 \setminus D_0$, we have proved that the considered simplification is compatible.

Assuming that $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification is compatible we get that (1) is perfect just by repeating the previous reasoning in the reverse direction; it is possible because Lemma 2 is an equivalence. □

Corollary. Let subsets $B_0, B_1, B_2, \ldots, B_{n+1}$ of K be such that

1. $\bigcup_{i=0}^{n+1} B_i = K$,
2. $B_0, B_1, B_2, \ldots, B_{n+1}$ meets the running intersection property,
3. for all $i = 1, \ldots, n+1$ sets $B_i \setminus (B_0 \cup \ldots \cup B_{i-1}) \neq \emptyset$.

Then defining for all $i = 0, 1, \ldots, n$

- $E_i = \bigcup_{j=0}^{n-i} B_j$,
- $D_i = \bigcup_{j=0}^{n+1-i} B_j \setminus B_{n+1-i}$,

the $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification is compatible for any distribution $\kappa(K)$.

Proof. It is easy to verify that sets E_i, D_i are defined in the way that

$$D_n \subset E_n \subsetneq E_{n-1}, D_{n-1} \subset E_{n-1} \subsetneq E_{n-2}, \ldots, D_1 \subset E_1 \subsetneq E_0, D_0 \subset E_0 \subsetneq K.$$

and

$$E_n = B_0, E_{n-1} \setminus D_n = B_1, E_{n-2} \setminus D_{n-1} = B_2, \ldots, E_0 \setminus D_1 = B_n, K \setminus D_0 = B_{n+1}.$$

Theorem 2 says that *any* sequence of pairwise consistent distributions which are defined for sets of variables meeting the running intersection property is perfect. Sequence (1) consists of marginals of κ, therefore all its elements are pairwise consistent and thus sequence (1) is perfect. This gives us that, due to the preceding Theorem, the considered simplification is compatible. □

5. CONCLUSIONS

We have presented a notion of a compatible $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification of a distribution $\kappa(K)$ and showed that a general $(E_0, D_0; \ldots; E_n, D_n)$-simplification is compatible if and only if all the distributions from the system (1) are also marginal to the resulting distribution

$$\bar{\kappa}^n = \kappa^{\downarrow E_n} \triangleright \kappa^{\downarrow E_{n-1} \setminus D_n} \triangleright \kappa^{\downarrow E_{n-2} \setminus D_{n-1}} \triangleright \ldots \triangleright \kappa^{\downarrow E_1 \setminus D_2} \triangleright \kappa^{\downarrow E_0 \setminus D_1} \triangleright \kappa^{\downarrow K \setminus D_0}.$$

The corollary presented at the end of the paper introduces a non-surprising fact that if a $(E_0, D_0; E_1, D_1; \ldots; E_n, D_n)$-simplification results in a decomposable model then it is also compatible.

ACKNOWLEDGEMENT

This research was partially supported by Grant Agency of the Academy of Sciences of the Czech Republic under grant A 2075302 and by the Ministry of Education, Youth and Sports of the Czech Republic under projects 1M0572 and 2C06019.

(Received August 10, 2006.)

REFERENCES

[1] R. Jiroušek: Composition of probability measures on finite spaces. In: Proc. 13th Conf. Uncertainty in Artificial Intelligence UAI'97 (D. Geiger and P. P. Shenoy, eds.), Morgan Kaufmann, San Francisco 1997, pp. 274–281.

[2] R. Jiroušek: Marginalization in composed probabilistic models. In: Proc. 16th Conf. Uncertainty in Artificial Intelligence UAI'00 (C. Boutilier and M. Goldszmidt, eds.), Morgan Kaufmann, San Francisco 2000, pp. 301–308.

[3] A. Perez: Information, ε-sufficiency and data reduction problems. Kybernetika *1* (1965), 297–323.

[4] A. Perez: Information theory methods in reducing complex decision problems. In: Trans. 4th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (J. Kožešník, ed.), Academia, Prague 1965, pp. 55–87.

[5] A. Perez: ε-admissible simplification of the dependence structure of a set of random variables. Kybernetika *13* (1977), 439–449.

[6] P. Somol, J. Novovičová, and P. Pudil: Notes on the evolution of feature selection methodology. Kybernetika *43* (2007), 713–730.

Radim Jiroušek, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: radim@utia.cas.cz