Kybernetika

Otakar Kříž

Conditional problem for objective probability

Kybernetika, Vol. 34 (1998), No. 1, [27]--40

Persistent URL: http://dml.cz/dmlcz/135183

KYBERNETIKA — VOLUME 34 (1998), NUMBER 1, PAGES 27–40

CONDITIONAL PROBLEM FOR OBJECTIVE PROBABILITY

OTAKAR KŘÍŽ

Marginal problem (see [5]) consists in finding a *joint distribution* whose *marginals* are equal to the given less-dimensional distributions. Let's generalize the problem so that there are given not only less-dimensional distributions but also *conditional probabilities*.

It is necessary to distinguish between objective (Kolmogorov) probability and subjective (de Finetti) approach ([4, 10]). In the latter, the *coherence problem* incorporates both probabilities and conditional probabilities in a unified framework. Different algorithms available for its solution are described e.g. in ([3, 4, 11]). In the context of the former approach, it will be shown that it is possible to split the task into solving the marginal problem independently and to subsequent solving pure "conditional" problem as certain type of optimization. First, an algorithm (*Conditional problem*) that generates a distribution whose conditional probabilities are equal to the given ones is presented. Due to the multimodality of the criterion function, the algorithm is only heuristical. Due to the computational complexity, it is efficient for small size problems e.g. 5 dichotomical variables.

Second, a method is mentioned how to unite *marginal* and *conditional* problem to a more general *consistency* problem for objective probability. Due to computational complexity, both algorithms are effective only for limited number of variables and conditionals. The described approach makes possible to integrate in the solution of the consistency problem additional knowledge contained e.g. in an empirical distribution.

1. INTRODUCTION

In *expert systems* with probabilistic background, there is a standard procedure to reconstruct a *joint distribution* from different *pieces of knowledge*. This joint distribution is then used as a tool to generate *aposteriori probabilities* of predicted variables given some fixed values other variables take. Then the alternative with the greatest value of aposteriori probability is a natural solution to the *decision making problem*. However, the first part of the process (i.e. to find such a *joint distribution* or at least to discover its potential existence) is referred to as *marginal problem* (see [5]). The ground for it is that the standard way of supplying *pieces of knowledge* is to use less-dimensional distributions that may be looked upon as the *marginals* of the theoretical *joint* one. The problem has been given certain attention especially for distributions on small *finite sets* recently. In general, there is a difference in algorithmical treatement of the problem for *objective* and *subjective* (de Finetti) *Bayesian approach*. The essence of the latter can be seen e.g. in [3, 4, 10, 11]. As far as the former approach is concerned: Sometimes, the IPFP algorithm (see [2]) is used to solve the marginal problem. In [1], it is suggested to use Lagrange multipliers to respect constraints. In [8], there is an algorithm presented (*Tetris*) checking if a set of marginals is consistent (and in positive case supplying one of feasible solutions). In [9], a modification of this algorithm (*Pentis*) is given that looks for the "best" approximation in l_1 norm even if the marginals are not consistent.

The main topic of this paper is to integrate a fixed set of *marginals* (original marginal problem) and a set of fixed *conditional probabilities* for *objective Bayesian approach*. Searching for different objects can be done in a "hard" or in a "soft" way. "Hard" stands for constraints when we try to solve sets of linear equations for equality constraints. "Soft" way means we try to enforce the results by minimization of a functional. If the constraints are contradictory, there exists no solution. One possible way to solve conditional problem is to convert it to a marginal problem by turning explicitely given conditional probabilities into additional linear constraints and to solve it then in a a standard way e.g. by using some package for linear programming. Let us remark that this incorporating equations with conditionals changes the character of the matrix, therefore more specific methods like [8] cannot be used and we have to apply integer programming or general linear programming methods. In this paper, we shall use another way namely minimizing a criterion function. Justification for this computationally more difficult approach is that thus we create a basis for solving more complex tasks. Namely, besides incorporating just marginals and conditionals, we may integrate the knowledge coming from an empirical distribution as the algorithm is using a "starting" distribution X^0. In this sense we construct a non-orthogonal projection of an empirical distribution on a set of all distributions complying with the marginals and at the same time we try to enforce fixed conditionals. And even more, we may give some weights to conditionals and reflect thus their relative precision or significance. Finally, using this approach, it is possible to include in the criterion function a term that would not only project X_0 but would force out a solution with e.g. higher entropy.

The basic idea is to look for a solution of the "marginal" part of the problem. It can be done with the help of linear programming or via e.g. *Tetris* algorithm [8]). Then using the parametrization "invariant moves" ([6]), preserving the marginals, we shall try to satisfy the conditions for conditional probabilities in the optimization procedure similar to one suggested in [7]. The outline of the paper is the following one (with underlined words serving as subsections titles):

The *consistency problem* $\mathrm{CP}(\mathcal{K}, W)$ (i.e. search for a joint distribution P with given *marginals* $\mathcal{K}$ and given *conditionals* W) can be defined as an optimization on the space of distributions.

The constraints are given by a set $\mathcal{K}$ of marginals and the criterion function Φ can be chosen as a distance $\Phi(P, W)$ between a distribution P and a (set of distributions having the same) fixed set W of conditionals. The extrem we look for is the minimum. Then, the *consistency problem* $\mathrm{CP}(\mathcal{K}, W)$ is solved if P is found for which the equality $\Phi(P, W) = 0$ holds.

The constraints $\mathcal{K}$ are automatically respected if

1. we found at least one distribution P_0 that has the marginals equal to the given ones and
2. changes of P_0 are done not in an arbitrary way but only in certain directions – the so-called *invariant moves.*

In other words, the joint distribution P_0 and *invariant moves* $V(\mathcal{K})$ (depending on $\mathcal{K}$) represent an efficient parametrization of the set $\mathcal{P}(\mathcal{K})$ of all joint distributions consistent with constraints $\mathcal{K}$.

Next, the change in distance function $\Phi(P,W)$ will be expressed as a function of the set V of *invariant moves.*

Then, the algorithm Conditional problem will be presented as a modification of an optimization program from [7] with the above mentioned specific distance function $\Phi(P,W)$.

If *invariant moves* $V(\mathcal{K})$ correspond to $\mathcal{K} = \emptyset$, P_0 can be any distribution e. g. the uniform one and we solve the pure *conditional problem.*

If $\mathcal{K} \neq \emptyset$, the starting distribution P_0 must comply with constraints $\mathcal{K}$, the corresponding *invariant moves* $V(\mathcal{K})$ preserve the value of the marginals and we solve general *consistency problem.* Formally, there is no difference in the form of the algorithm for both cases.

2. OPTIMIZATION ON DISTRIBUTIONS

Given a set $\mathcal{K}$ of less dimensional distributions

$$\mathcal{K} = \{P_{S_1}, P_{S_2} \dots P_{S_t}\}$$

(S_k denotes a set of variables whose behaviour P_{S_k} describes) and an arbitrary functional Φ on the space $\mathcal{P}$ of all joint distributions, find such a representative $\widehat{P}_{\xi_1\xi_2\dots\xi_n}$ from the class $\mathcal{P}(\mathcal{K})$ of all joint distributions $P_{\xi_1\xi_2\dots\xi_n}$ consistent with a set $\mathcal{K} = \{P_{S_1}, P_{S_2} \dots P_{S_t}\}$ for which the value $\Phi\left(\widehat{P}_{\xi_1\xi_2\dots\xi_n}\right)$ of the functional Φ achieves the extremal value:

$$\widehat{P}_{\xi_1\xi_2\dots\xi_n} = \underset{P\in\mathcal{P}(\mathcal{K})}{\arg\operatorname{ext}}\ \Phi(P)$$

where

$$\mathcal{P}(\mathcal{K}) = \left\{P \in \mathcal{P} \mid P^{S_i} = P_{S_i}\ \ i = 1, 2, \dots \mathrm{card}(\mathcal{K})\right\}.$$

An upper index S_i applied to symbol P means marginalization.

3. CONSTRAINTS

Constraints in the formulation of the problem can be expressed formally by the matrix equation $\boldsymbol{Ax} = \boldsymbol{b}$ where $\boldsymbol{x}$ is vector of ordered values of the joint distribution P on k atoms of the algebra of subsets of a basic finite set U. The vector $\boldsymbol{b}$ contains the values of *all* small-dimensional distributions that are considered as fixed and the matrix $\boldsymbol{A}$, referred to as *incidence matrix*, consists of zeroes and ones only. The interpretation (in terms of original problem) is the following one: If $a_{ij} = 1$

then the probability value x_j of the j-th atom (of sample space on which the joint distribution is defined) contributes to the probability value b_i on respective atom i (of some of the marginals from which vector $\boldsymbol{b}$ is composed). The equation $\boldsymbol{Ax} = \boldsymbol{b}$ describes exactly the "marginalization" conditions for some selected marginal (less-dimensional) distributions (that are postulated as the knowledge base $\mathcal{K}$).
In ([6]), a special type of parametrization of $\boldsymbol{x}$ (preserving the constraints $\boldsymbol{Ax} = \boldsymbol{b}$) was introduced under the name of *invariant moves.* (It is such a transformation of $\boldsymbol{x}$ that when we "move" the values of $\boldsymbol{x}$, the values of $\boldsymbol{b}$ remain "invariant".)
Invariant moves are given by partitions $\{P, N, U - (P \cup N)\}$ of the set U of k points (on which a joint distribution is defined). The groups P and N are determined with the help of the matrix $\boldsymbol{A}(m, k)$ by m following conditions:

$$\sum_{j \in P} a_{ij} = \sum_{j \in N} a_{ij} \ for\ i = 1, 2, \ldots m.$$

If we increase values of distribution in all points in the P-group by a real number β and simultaneously decrease by the number β the values in points of the N-group, all marginals (described by $\boldsymbol{A}$) remain unchanged. The incidence matrix $\boldsymbol{A}$ can degenerate into a "row" vector consisting of all "ones". This corresponds to the equation $\boldsymbol{Ax} = \boldsymbol{b}$ where vector $\boldsymbol{b}$ has degenerated to the scalar 1. Then, $\mathcal{K} = \emptyset$, $\mathcal{P}(\mathcal{K}) = \mathcal{P}_k$ (i.e. all distributions on k elements) and invariant moves for this simplex $\mathcal{P}_k$ include all $\binom{k}{2}$ pairs with P and N groups containing always one atom only.

4. CRITERION FUNCTION

Given a fixed set W of conditional probabilities, we look for a joint distribution X having its conditionals (on corresponding definition sets) equal to the ones in W. If there are more solutions, we wish to construct one of them. If there is no one, we accept a distribution X where the "distance" $\Phi(X, W)$ between W and the conditionals of X is the smallest one in a sense. In the sequel, $\Phi(X, W)$ will be selected as the sum of absolute values of differences of respective conditionals.

1. Let U denote a finite set. It's elements will be denoted as u. Considered as one-elements sets $\{u\}$, they are atoms of algebra of subsets of U.
2. Let X be a probability distribution on U. $X(u)$ denotes value of probability X for the respective atom $\{u\}$.
3. Let w denote a *conditional* $(C_w, D_w, P_{C_w|D_w}) \in 2^U \times 2^U \times \langle 0, 1\rangle$ where two sets C_w, D_w are subsets of U (i.e. $C_w \in U$, $D_w \in U$) and the number $P_{C_w|D_w}$ fulfills some requirements:
   (a) If $D_w = \emptyset$, then $P_{C_w|D_w}$ can be any number from $\langle 0, 1\rangle$.
   (b) If $C_w = \emptyset$, then $P_{C_w|D_w}$ must be 0.
   (c) If $C_w \cap D_w = \emptyset$ & $D_w \neq \emptyset$, then $P_{C_w|D_w}$ must be 0.
   (d) If $C_w \supseteq D_w$, then $C_w \cap D_w = D_w$ and $P_{C_w|D_w}$ must be 1.

4. Let cc_w, dd_w be the least integers expressing $P_{C_w|D_w}$ as rational number i.e.

$$P_{C_w|D_w} = \frac{cc_w}{dd_w}.$$

Set of all fixed *conditionals* w will be denoted as W.

5. Let L denote the least integer number such that all $LX(u)$ are integers and L is also divisible by denominators dd_w for all $w \in W$. Let $cd_w(X)$ denote L times the probability $X(C_w D_w)$ of intersection $C_w D_w$, and $d_w(X)$ denote L times the probability $X(D_w)$. Conditional probability $X(C_w|D_w)$ of C_w given D_w if X is the joint probability on U is then given by

$$X(C_w|D_w) = \frac{X(C_w D_w)}{X(D_w)} = \frac{cd_w(X)}{d_w(X)}$$

where integer numbers $cd_w(X), d_w(X)$ may have common divisor unlike to cc_w and dd_w.

6. Let $V(\boldsymbol{A})$ be a set of all invariant moves for incidence matrix $\boldsymbol{A}$. Its elements are denoted as v and P_v, N_v are *positive* and *negative* sets of the move v. Remark: The symbol P is used in two senses: probability and positive. P_v is a subset of U, P is a joint distribution, $P_{C_w|D_w}$ is a number in a *conditional* w and $P(C_w|D_w)$ is a conditional probability of C_w given D_w if P is the joint probability on U.

7. The "distance" $\Phi(X, W)$ of probability X from set W of conditionals (all defined on U) is given by

$$\Phi(X, W) = \sum_{w \in W} \left| \left[\frac{cc_w}{dd_w} - \frac{cd_w(X)}{d_w(X)} \right] \right|.$$

If $\Phi(X, W) = 0$, the probability X is consistent with the given set W of conditionals i.e.

$$\forall_{w \in W} \;\; X(C_w|D_w) = P_{C_w|D_w}.$$

8. Let us suppose that the distribution X changed to distribution X_{kv} using the move v and performing k steps in direction v. The transformed X_{kv} is given by formula

$$\begin{aligned} X_{kv}(u) &= X(u) + k && u \in P_v \\ &= X(u) - k && u \in N_v \\ &= X(u) && u \in U - P_v \cup N_v. \end{aligned}$$

Then we get another "distance"

$$\Phi(X_{kv}, W) = \sum_{w \in W} \left| \left[\frac{cc_w}{dd_w} - \frac{cd_w(X_{kv})}{d_w(X_{kv})} \right] \right|$$

where the "integer" (due to multiplying by L) probabilities $cd_w(X_{kv})$ and $d_w(X_{kv})$ can be expressed as

$$cd_w(X_{kv}) = \sum_{u \in C_w D_w P_v} (LX(u)+k) + \sum_{u \in C_w D_w N_v} (LX(u)-k) + \sum_{u \in C_w D_w (\overline{P}_v \cap \overline{N}_v)} LX(u)$$

and

$$d_w(X_{kv}) = \sum_{u \in D_w P_v} (LX(u)+k) + \sum_{u \in D_w N_v} (LX(u)-k) + \sum_{u \in D_w (\overline{P}_v \cap \overline{N}_v)} LX(u).$$

These expressions may be modified to

$$cd_w(X_{kv}) = cd_w(X) + k \left[\sum_{u \in C_w D_w P_v} 1 - \sum_{u \in C_w D_w N_v} 1 \right]$$

$$d_w(X_{kv}) = d_w(X) + k \left[\sum_{u \in D_w P_v} 1 - \sum_{u \in D_w N_v} 1 \right].$$

If we introduce following abbreviations for cardinalities of respective sets

$$cd_{wv} = |C_w D_w P_v| - |C_w D_w N_v|$$

and

$$d_{wv} = |D_w P_v| - |D_w N_v|$$

the formula for $\Phi(X_{kv}, W)$ can be transcribed into form

$$\Phi(X_{kv}, W) = \sum_{w \in W} \left| \left[\frac{cc_w}{dd_w} - \frac{cd_w(X) + k \cdot cd_{wv}}{d_w(X) + k \cdot d_{wv}} \right] \right|.$$

Let us stress that the integers cd_{wv} and d_{wv} can be calculated for all combinations from $W \times V$ once for all before the optimizing cycle. Some of them may be zero. The formula implies that it is possible to look for minimum of $\Phi(X_{kv}, W)$ changing only the value of step k.

9. Let $k_{\max}(v, X^l)$ denote the value of k recommended for the final transformation of distribution X^l into the distribution X^l_{kv}. Then, $k_{\max}(v, X^l)$ either yields the best improvement of distance criterion in the direction v or it is limited by a free parameter k_0 i.e

$$k_{\max}(v, X^l) = \min \left\{ \operatorname*{arg\,min}_{k \in N \cup (-N)} \Phi(X^l_{kv}, W),\ k_0 \right\}.$$

10. Starting from a given (possibly arbitrary) joint probability X^0, let us construct a sequence X^0,X^1,X^2,... such that X^l is generated from X^{l-1} applying the move v where $v = \operatorname{mod}(l, |V|) + 1$.
The stopping rule may be of the form $\Phi(X^l, W) = 0$ or $l \geq l_0$ for a fixed l_0.

In subsequent steps, only k is changing in evaluating $k_{\max}(v, X^l)$ and then, after the $k_{\max}(v, X^l)$ was fixed, we perform a transformation

$$X^l \longrightarrow X^l_{k_{\max}v} = X^{l+1}$$

where changes of X^l are performed only in points $u \in P_v \cup N_v$ inciding with the move v and not everywhere in U and the numbers $cd_w(X^{l+1})$ and $d_w(X^{l+1})$ must be updated accordingly.

5. ALGORITHM: CONDITIONAL PROBLEM

The type of extremalization is selected as *minimum*. Concrete type of function Φ is $\Phi(X, W)$.

There are four cycles in the algorithm:

1. *set_of_moves cycle* determines that optimization over all $|V|$ *invariant moves* is to be performed y times (y is a free parameter).
2. *single_move cycle*: optimization is performed for each *invariant move* v from $V(\boldsymbol{A})$. On request (by parameter o), the order of running through the moves set can be reversed.
3. *positive* or *negative directions*: prolongation of tentative calculation of the functional as long as there exists a decrease in functional value from the previous step. *Positive direction* means that values from P are incremented by *positive* value of the module k in each step. The prolongation (of this linear search) can be constrained from above by a (free) parameter k_0.

The following description of the algorithm is a very rough one, featuring some important points and hints only.

1. enter the starting distribution $\boldsymbol{X}^0$
2. select L to have sufficiently small step

$$L = \text{l.c.p}\left(\{dd_w\}_{w \in W} \bigcup \frac{1}{\text{g.c.d.}(\{X^0(u)\}_{u \in U})}\right)$$

   where *l.c.p* stands for the *least common product* and *g.c.d* stands for the *greatest common divisor.*

3. enter the set $V(\boldsymbol{A})$ of all *invariant moves*
4. enter the set W of all conditionals $\left(C_w, D_w, \frac{cc_w}{dd_w}\right)$
5. calculate the matrices cd_{wv} and d_{wv} for $w \in W$, $v \in V$.
6. set-of-moves cycle: duration according to free parameter y. It makes further possible "reoptimization" of the final X
7. calculate value of $\Phi(X, W)$ in the given "point" (distribution) X
8. single-invariant move cycle: order given by parameter o
9. new-move: select *invariant move* v

10. proceed in *positive direction* as long as the functional decreases or unless the prolongation is inhibited by the parameter k_0 or unless Φ achieves the boundary 0.0 or unless $LX(u) - k$ achieves the zero for an atom $u \in N_v$.
11. proceed in *negative direction* as long as the functional decreases or unless the prolongation is inhibited by the parameter k_0 or unless the boundary 0.0 is achieved or unless $LX(u) - k$ achieves the zero for an atom $u \in P_v$.
12. select the direction with better decrease in criterial functional. If there is no such direction increase the single_invariant_moves cycle index and proceed to new-move:.
    If there existed possible improvement, adapt the vector X by the changes in all u_i from the P_v and N_v group of the actual *invariant move* (P_v, N_v), recalculate $cd_w(X), d_w(x)$ for $u \in P_v \cup N_v$.
13. If $\Phi(X, W) = 0$ go to *end:*, otherwise increase the single_invariant_moves cycle index and proceed to *new-move:*.
14. end of single-invariant moves cycle:
15. end of set-of-moves cycle:
16. end: printing and storing of results.

6. EXAMPLE

Let us consider three dichotomic variables ξ_1, ξ_2, ξ_3. They take values from {0,1}. It means that there are 2^3 atoms, i.e. $U = \{1,2,3,4,5,6,7,8\}$. We suppose that $\mathcal{K} = \emptyset$ i.e. we have not fixed any marginal and the only requirement is that the output of the algorithm is a distribution, if input was a distribution. Then, the corresponding matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = (1\,1\,1\,1\,1\,1\,1\,1).$$

For this specific matrix $\boldsymbol{A}$, the set $V(\boldsymbol{A})$ of all *invariant moves* can be derived directly and consists of all 28 pairs of 8 elements of U. Then, $V(\boldsymbol{A})$ is

$$V(\boldsymbol{A}) = \{(1,2),(1,3),\dots(6,8),(7,8)\}.$$

Let us suppose our requirement is that two conditional probabilities are to be fixed, namely $P_{\xi_1|\xi_2}(0|0) = 1/8$ and $P_{\xi_2|\xi_1\xi_3}(0|1,0) = 4/13$. In our previous terminology, the set W of conditionals has two elements and it is defined by

$$W = \{(\langle 10101010\rangle,\ \langle 11001100\rangle,\ 1/8),\ (\langle 11001100\rangle,\ \langle 01010000\rangle,\ 4/13)\}.$$

The sets are given by their characteristic functions, e.g. $D_{w_1} = \langle 11001100\rangle$ means that the set D_{w_1} contains atoms 1,2,5,6. The requirements (definition of W) are written in certain pseudo-language. We shall keep as comments (line starting with c) original values of conditional probabilities for the starting distribution X^0 to ease up comparing the results.

```
c-----(1)/(2):1=<4 1>/5
1/8       <10101010|                    <<<<<<<
           11001100>
c4/5      <10101010|
c          11001100>
c1/5      <01010101|
c          11001100>
c-----(2)/(1,3):2=<1 10>/11
4/13       <11001100|                   <<<<<<<<<
            01010000>
c1/11      <11001100|
c           01010000>
c10/11     <00110011|
c           01010000>
```

In the sequel, we shall present behaviour of the algorithm *Conditional problem* when it is applied to two different distributions.

First, starting distribution X^0 was selected as

```
(1,2,3)=<4/20 1/20 5/20 10/20 0/20 0/20 0/20 0/20>
```

The interpretation of the previous line is e. g. $P_{\xi_1\xi_2\xi_3}(0,1,0) = 5/20$. The marginal distributions of X^0 are

```
 (1)=<9/20 11/20>
 (2)=<5/20 15/20>
 (3)=<20/20 0/20>
 (1,2)=<4/20 1/20 5/20 10/20>
 (1,3)=<9/20 11/20 0/20 0/20>
 (2,3)=<5/20 15/20 0/20 0/20>
```

its conditional probabilities are

```
 (1)/(2):1=<4 1>/5         <<<<<<<<<<<<<<<     change to 1/8
 (1)/(2):2=<1 2>/3
 (1)/(3):1=<9 11>/20
........................
 (2)/(1,3):2=<1 10>/11       <<<<<<<<<<<<<<<<   change to 4/13
 (2)/(1,3):3=<0 0>/0
 (2)/(1,3):4=<0 0>/0
....................
 (2,3)/(1):2=<1 10 0 0>/11
```

and in detail (with explicitely given sets C_w, D_w) for conditionals that we have to change

```
c-----(1)/(2):1=<4 1>/5
4/5       <10101010|         <<<<<<<<<<<<<<<<<<<<<<<<    change to 1/8
           11001100>
```

```
1/5       <01010101|
           11001100>
..................
c-----(2)/(1,3):2=<1 10>/11
1/11      <11001100|            <<<<<<<<<<<<<<<<<<<<<<    change to 4/13
           01010000>
10/11     <00110011|
           01010000>
.................................
```

The resulting joint distribution X^{35}

```
(1,2,3)=<23/520 20/520 1/520 45/520 0/520 141/520 0/520 290/520>
```

with marginals

```
(1)=<24/520 496/520>
(2)=<184/520 336/520>
(3)=<89/520 431/520>
(1,2)=<23/520 161/520 1/520 335/520>
(1,3)=<24/520 65/520 0/520 431/520>
(2,3)=<43/520 46/520 141/520 290/520>
```

and conditional probabilities for sets where we required the change

```
c-----(1)/(2):1=<1 7>/8
1/8       <10101010|                         <<<<<<<<<<
           11001100>
7/8       <01010101|
           11001100>
..................................
c-----(2)/(1,3):2=<4 9>/13
4/13      <11001100|                          <<<<<<<<<<
           01010000>
9/13      <00110011|
           01010000>
.......................................
```

The importance of different parameters (e. g. o or k_0) influencing application of different invariant moves can be seen from the following tables. In the first table, 8 joint distributions P^j are given that were offered by the algorithm as a solution of the conditional problem for different "mixing" parameters o and k_0. Integer numbers corresponding to different atoms x_l must be divided by the number from column "/" (i.e. 520) to obtain probabilities $P^j(\{x_l\})$

P^j / Atoms	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	/	O.K.
P^1	22	96	127	216	0	58	0	1	/520	*
P^2	18	27	130	61	0	99	86	99	/520	
P^3	20	26	17	59	0	114	100	184	/520	
P^4	23	20	1	45	0	141	1	289	/520	*
P^5	16	115	130	259	0	0	0	0	/520	
P^6	19	98	133	221	1	42	6	0	/520	
P^7	21	98	131	221	0	49	0	0	/520	*
P^8	23	20	1	45	0	141	0	290	/520	*

Unfortunately, not every joint distribution is the exact solution of the conditional problem. Only joint distributions P^j denoted by $*$ in column O.K. of the first table fulfill exactly requirement for conditional probabilities. Some P^j are just "approximate" solutions as can be deduced from the second table using columns $\stackrel{?}{=}$ and Err_1 and Err_2. But even for not exact solutions, the errors are very small.

P^j	o	k_0	$P^j_{\xi_1\|\xi_2}(0\|0)$	$\stackrel{?}{=}$	$\frac{cc_1}{dd_1}$	Err_1	$P^j_{\xi_2\|\xi_2\xi_3}(0\|1,0)$	$\stackrel{?}{=}$	$\frac{cc_2}{dd_2}$	Err_2
P^1	1	1	$\frac{22}{176}$	$=$	$\frac{1}{8}$	0	$\frac{96}{312}$	$=$	$\frac{4}{13}$	0
P^2	1	50	$\frac{18}{144}$	$=$	$\frac{1}{8}$	0	$\frac{27}{88}$	$\neq$	$\frac{4}{13}$	0.0008
P^3	1	100	$\frac{20}{160}$	$=$	$\frac{1}{8}$	0	$\frac{26}{85}$	$\neq$	$\frac{4}{13}$	0.0018
P^4	1	200	$\frac{23}{184}$	$=$	$\frac{1}{8}$	0	$\frac{20}{65}$	$=$	$\frac{4}{13}$	0
P^5	0	5000	$\frac{16}{131}$	$\neq$	$\frac{1}{8}$	0.002	$\frac{115}{374}$	$\neq$	$\frac{4}{13}$	0.0002
P^6	0	3	$\frac{21}{168}$	$=$	$\frac{1}{8}$	0	$\frac{98}{319}$	$\neq$	$\frac{4}{13}$	0.0041
P^8	1	25	$\frac{23}{184}$	$=$	$\frac{1}{8}$	0	$\frac{20}{65}$	$=$	$\frac{4}{13}$	0

Second, let us suppose that the starting distribution X^0 is uniform

```
(1,2,3)=<1/8 1/8 1/8 1/8 1/8 1/8 1/8 1/8>
```

with conditional probabilities

```
c-----(1)/(2):1=<1 1>/2
1/2       <10101010|                    <<<<<<<<<<<<<<<
           11001100>
1/2       <01010101|
           11001100>
.............................
c-----(2)/(1,3):2=<1 1>/2
1/2       <11001100|                         <<<<<<<<<<<<<
           01010000>
1/2       <00110011|
           01010000>
```

Then after applying the algorithm Conditional problem to X^0, there were successively 26 *moves* used, with limiting condition that 5 steps (as $k_0 = 5$ was set) can be made at most in each direction (*move*) in one big cycle. Then, the resulting probability X^{26} is

```
(1,2,3)=<0/104 8/104 18/104 18/104 5/104 27/104 15/104 13/104>
```

with marginals

```
(1)=<38/104 66/104>
(2)=<40/104 64/104>
(3)=<44/104 60/104>
(1,2)=<5/104 35/104 33/104 31/104>
(1,3)=<18/104 26/104 20/104 40/104>
(2,3)=<8/104 36/104 32/104 28/104>
```

and corresponding conditional probabilities

```
c-----(1)/(2):1=<1 7>/8
1/8       <10101010|                    <<<<<<<<<<<<<
           11001100>
7/8       <01010101|
           11001100>
..............................
c-----(2)/(1,3):2=<4 9>/13
4/13      <11001100|                      <<<<<<<<<<<<<
           01010000>
9/13      <00110011|
           01010000>
```

As the starting distribution X^0 was uniform, the resulting distribution X^{26} should have the greatest entropy among all distributions with the given conditionals.

7. CONCLUSIONS

1. It is possible to solve *consistency problem* $\mathrm{CP}(\mathcal{K}, W)$ or a pure *conditional problem* $\mathrm{CP}(\emptyset, W)$ in form of optimization on space of distributions.

2. Due to multimodal character of the criterion function $\Phi(X, W)$ and different starting distributions X^0, the algorithm can finish in a distribution X^l for which $\Phi(X, W) > 0$, even if the problem $\mathrm{CP}(\mathcal{K}, W)$ or $\mathrm{CP}(\emptyset, W)$ has a solution.
3. In these situations, changing the steering parameters of the optimization algorithm (e. g. order of applying invariant moves V, limiting length k_0 of search in each direction (*move*)) can find another distribution X^l for which $\Phi(X, W) = 0$.
4. If we are interested in strict "yes" or "no" reply to consistency, it may be better to use linear programs.
5. If we wish to integrate the knowledge from an empirical distribution, it should be used as a starting distribution X^0 in the algorithm. This way we can get one of possible (non-orthogonal) projections of X^0 on set of distributions with the given marginals and conditionals.
6. The algorithm is efficient only for small-size problems e. g. joint distributions for 5 dichotomical variables.
7. Due to treating the problem in integers, rounding errors in evaluating criterion are avoided.
8. There are no specific requirements (e. g. complying with a graphical model) on the structure of the joint distribution.

(Received November 7, 1997.)

REFERENCES

[1] P. Cheeseman: A method of computing generalized Bayesian probability values of expert systems. In: Proceedings of the 6-th Joint Conference on Artificial Intelligence (IJCAI-83), Karlsruhe, pp. 198–202.

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

[3] A. Gilio and S. Ingrassia: Geometrical aspects in checking coherence of probability assessments. In: IPMU'96: Proceedings of the 6th International IPMU Conference (B. Bouchon–Meunier, M. Delgado, J. L. Verdegay, M. A. Vila, R. Yager, eds.), Granada 1996, pp. 55–59.

[4] G. Coletti and R. Scozzafava: Characterization of coherent conditional probabilities as a tool for their assessment and extension. Internat. J. Uncertainty, Fuzziness and Knowledge–Based Systems, *4* (1996), 2, 103–127.

[5] H. G. Kellerer: Verteilungsfunktionen mit gegebenen Marginalverteilungen. Z. Wahrsch. verw. Gebiete *3* (1964), 247–270.

[6] O. Kříž: Invariant moves for constructing extensions of marginals. In: IPMU'94: Proceedings of the 5th International IPMU Conference (B. Bouchon–Meunier, R. Yager, eds.), Paris 1994, pp. 984–989.

[7] O. Kříž: Optimizations on finite–dimensional distributions with fixed marginals. In: WUPES 94: Proceedings of the 3-rd Workshop on Uncertainty Processing (R. Jiroušek, ed.), Třešť 1994, pp. 143–156.

[8] O. Kříž: Marginal problem on finite sets. In: IPMU'96: Proceedings of the 6-th International IPMU Conference (B. Bouchon-Meunier, M. Delgado, J. L. Verdegay, M. A. Vila, R. Yager, eds.), Granada 1996, Vol. II, pp. 763–768.

[9] O. Kříž: Inconsistent marginal problem on finite sets. In: Distributions with Given Marginals and Moment Problems (J. Štěpán, V. Beneš, eds.), Kluwer Academic Publishers, Dordrecht – Boston – London 1997, pp. 235–242.

[10] R. Scozzafava: A probabilistic background for the management of uncertainty in Artificial Intelligence. European J. Engineering Education *20* (1995), 3, 353–363.

[11] P. Vicig: An algorithm for imprecise conditional probability assesment in expert systems. In: IPMU'96: Proceedings of the 6-th International IPMU Conference (B. Bouchon-Meunier, M. Delgado, J. L. Verdegay, M. A. Vila, R. Yager, eds.), Granada, 1996, Vol. I, pp. 61–66.

Ing. Otakar Kříž, CSc., 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: kriz@praha.ud.cez.cz