Perfect Sequences: A Contribution to Structuring Conditional Independence Models

Gernot D. Kleiter
Department of Psychology
University of Salzburg
Austria
gernot.kleiter@sbg.ac.at

Radim Jiroušek
Department of Decision-Making Theory
Inst. of Inform. Theory and Automation
Czech Republic
radim@utia.cas.cz

Abstract

The representation of conditional independence models by perfect sequences provides an alternative to Bayesian networks and essential graphs. The paper discusses properties of perfect sequences that are relevant with respect to different structures of conditional independence models. Boundary variables (related to terminal nodes in a directed graph representation) are used to find the number of labeled and unlabeled models and to enumerate parts of the model space. Structuring principles are further applied to the evaluation of whole conditional independence models in learning models from data.

1 Introduction

Conditional independence (CI) models build a rich class of models that describe probabilistic dependences and independences in complex uncertain domains. They are defined for a finite set of variables $\{X_i\}_{i\in N}$; in this paper we only consider variables X_i with finite sets of values $\mathbf{X}_i$. Probability distributions of these variables are denoted by Greek characters: $\pi, \kappa, ...$ (with indices, if necessary). Subsets of variables $(X_i)_{i\in K}$ (for $K \subseteq N$) are denoted simply by X_K. Correspondingly, x_K denote a combination of the values of the variables X_K (a vector from a corresponding Cartesian product set):

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

Thus, for example, a $|K|$-dimensional probability distribution is denoted by $\pi(x_K)$, and $\pi(x_M)$ is (for $M \subset K$) its marginal distribution. Notice that when considering marginal distributions $\pi(x_M)$ of $\pi(x_K)$ we do not exclude situations when $M = \emptyset$; in this case $\pi(x_\emptyset) = 1$.

The structure of a CI model is specified by a set of triplets $X_{K_1} \perp\!\!\!\perp X_{K_2} | X_{K_3}$, where K_1, K_2, and K_3 denote disjoint subsets of N. Each triplet denotes the conditional independence of X_{K_1} and X_{K_2} given X_{K_3}, that is, $\pi(x_{K_1}, x_{K_2}|x_{K_3}) = \pi(x_{K_1}|x_{K_3})\ \pi(x_{K_2}|x_{K_3})$.

The best known subclass of CI models is that of *Bayesian networks*. They represent the structure of CI models by directed acyclic graphs. Bayesian networks, though, over-specify CI models because many models can be represented by different but probabilistically equivalent graphs. *Essential graphs* [1] avoid this problem; they represent the structure of CI models by graphs with directed and undirected edges such that every model corresponds to one and only one graph. *Perfect sequences* are a non-graphical representation of CI models [4, 5]. Whereas Bayesian networks and essential graphs decompose the joint distribution into conditional distributions, perfect sequences represent the joint distribution by a composition of marginal distributions. The aim of the present paper is to point to a number of advantages offered by perfect sequences when they are applied to learning CI models from data.

2 Operator of composition and compositional models

Fundamental to perfect sequences is the following operator of composition (for a detailed discussion of the operator see [4, 5]):

Definition 1 (Composition) *For two arbitrary distributions $\pi(x_K)$ and $\kappa(x_L)$ their* composition *is given by the formula:*

$$\pi(x_K) \triangleright \kappa(x_L) = \begin{cases} \frac{\pi(x_K)\kappa(x_L)}{\kappa(x_{K\cap L})} & \text{if } \pi(x_{K\cap L}) \ll \kappa(x_{K\cap L}),^1 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

If K and L are disjoint, then $\kappa(x_{K\cap L}) = 1$ and the right hand side degenerates to a simple product $\pi \cdot \kappa$. Most often we will deal with consistent distributions which dominate each other.

Definition 2 (Pairwise consistent distributions) *Two distributions, $\pi(x_K)$ and $\kappa(x_L)$, are pairwise consistent iff*

$$\pi(x_{K\cap L}) = \kappa(x_{K\cap L}).$$

In this case, if for any $x_{K\cap L}$ $\pi(x_{K\cap L}) = 0$ (and therefore also $\kappa(x_{K\cap L}) = 0$), a product of two zeros occurs in the numerator and we take, quite naturally, $\frac{0\cdot 0}{0} = 0$.

[1] $\pi(x_{K\cap L}) \ll \kappa(x_{K\cap L})$ means that $\kappa(x_{K\cap L})$ dominates $\pi(x_{K\cap L})$. In the finite case, which is considered in the paper, this simplifies to the condition that for all $x_{K\cap L} \in \mathbf{X}_{K\cap L}$

$$\kappa(x_{K\cap L}) = 0 \Longrightarrow \pi(x_{K\cap L}) = 0.$$

The result of the composition of two distributions is given by the following theorem proven in [4].

Theorem 1 *If $\pi(x_K) \triangleright \kappa(x_L)$ is defined, then it is a probability distribution over $X_{K \cup L}$ and its marginal distribution of X_K equals π:*

$$(\pi \triangleright \kappa)(x_K) = \pi(x_K).$$

We next consider a system of n low-dimensional distributions $\pi_1(x_{K_1})$, ..., $\pi_n(x_{K_n})$. When the composition operator is applied iteratively to this sequence, it constructs multidimensional distribution. More precisely, the formula $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$, if it is defined, determines a joint distribution of the variables X_N (in case that $N = K_1 \cup K_2 \cup \ldots \cup K_n$). Because the operator $\triangleright$ is neither commutative nor associative, we apply the operators from left to right only, i.e.,

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \ldots \triangleright \pi_n = (\ldots((\pi_1 \triangleright \pi_2) \triangleright \pi_3) \triangleright \ldots \triangleright \pi_n).$$

Therefore, in order to construct a compositional model (that is, a multidimensional distribution, which is composed from low-dimensional ones) it is sufficient to determine a sequence – we will call it a *generating sequence* – of low-dimensional distributions. Thus, a generating sequence $\pi_1, \pi_2, \ldots, \pi_n$ defines the multidimensional distribution

$$\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n,$$

which is an $|N|$-dimensional distribution of over $\mathbf{X}_N$.

3 Perfect and efficient perfect sequences

The subclass of generating sequences most powerful to represent multidimensional distributions are *perfect sequences.*

Definition 3 (Perfect sequence) *A generating sequence of probability distributions $\pi_1, \pi_2, \ldots, \pi_n$ is called* perfect *if $\pi_1 \triangleright \ldots \triangleright \pi_n$ is defined and*

$$\begin{aligned}
&\pi_1 \triangleright \pi_2 = \pi_2 \triangleright \pi_1,\\
&\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_3 \triangleright (\pi_1 \triangleright \pi_2),\\
&\qquad \vdots\\
&\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n = \pi_n \triangleright (\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_{n-1}).
\end{aligned}$$

From this definition one can hardly see the importance of perfect sequences. This importance becomes clearer from the following characterization theorem (for its proof see [5]).

Theorem 2 *A sequence of distributions $\pi_1, \pi_2, \ldots, \pi_n$ is perfect* iff *all the distributions from this sequence are marginals of the distribution $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)$.*

In the following paragraphs we will deal also with properties of the structure of perfect sequences. Where appropriate we will point to the parallel properties in graph representations. For the properties to be discussed it is convenient to treat perfect sequences as lists of sets that obey certain "syntactical" rules. In this context we call K_i a *factor*. Since there is a bijection between indices and variables, it is convenient not to distinguish them. So we will speak, for example, of variables in a factor, and write $\{1\},\{2\},\{1,2,3\},\{3,4\}$ as a shorthand notation for $\pi_1(x_{\{1\}}) \triangleright \pi_2(x_{\{2\}}) \triangleright \pi_3(x_{\{1,2,3\}}) \triangleright \pi_4(x_{\{3,4\}})$.

A perfect sequence, or the corresponding sequence of factors, is an *efficient perfect sequence* (EP sequences), if it fulfills the subset condition.

Definition 4 (Efficient perfect sequence) *A perfect sequence $\pi_1(x_{K_1})$, ..., $\pi_n(x_{K_n})$. is an efficient perfect sequence, if the sequence K_1,, ... , K_n, fulfills the subset condition which holds if for all K_i and K_j, $1 \leq i < j \leq m$,*

$$(K_i \subset K_j) \Rightarrow \exists K_k (1 \leq k < j, k \neq i)(K_k \cap K_j \neq \emptyset)\,.$$

Example 1 The sequence $\{1\},\{2\},\{1,3\},\{1,4\},\{2,3,5\}$ does not fulfill the subset condition because $\{1\}$ is a subset of $\{1,3\}$ and there is no third factor left of $\{1,3\}$ with a non-empty intersection with it; $\{2\},\{1,3\},\{1,4\},\{2,3,5\}$, however, fulfills the subset condition.

Each factor appearing in a sequence consists of a head and a tail.

Definition 5 (Head and tail) *In each factor of a perfect sequence the set of variables that appears the first time (when the sequence is read from left to right) is called the "head" and the set of remaining variables is called the "tail" of the factor.*

Example 2 In the sequence $\{\overline{1}\},\{\overline{2}\},\{1,2,\overline{3}\},\{1,2,\overline{4}\},\{3,\overline{5}\}$ the heads are overlined.

We write the head on the right, the tail on the left hand side of each factor. The subset condition now can be expressed simply by saying that no factor on the left hand side is identical to a tail of another factor on the right hand side. An EP sequence in which all factors consist of heads only, corresponds to a partition of the N variables and each factor corresponds to an undirected completely connected component. If the sequence consists of only one factor, then the model does not contain any independence. If the sequence consists of factors having one variable only, then all variables are independent and it corresponds to the edgeless graph.

4 Boundary variables

In a directed graph a node is a terminal node if it has no children. In a Bayesian network (where the direction of arcs is not always fixed because of probabilistically equivalent models) a node may be a terminal node in one graph representation but not in another one. In a Bayesian network a node is a terminal

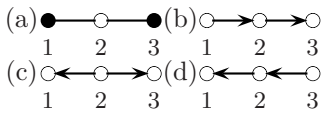

Figure 1: (a) Essential graph (corresponding EP sequence $S_a = \{1,2\},\{2,3\}$); the filled vertices 1 and 3 are boundary variables. (b), $S_b = \{1,2\},\{2,3\}$), (c), $S_c = \{2,1\},\{2,3\}$), and (d), $S_d = \{3,2\},\{2,1\}$) are the three probabilistically equivalent Bayesian networks of (a). In none of these three structures vertex 2 is a terminal node. It therefore is not a boundary variable and not filled in (a). 1 and 3 are terminal nodes in at least one of the Bayesian networks (b), (c) or (d). They therefore are marked as boundary variables in (a).

node—we will say a boundary variable in the present paper—if it admits at least one probabilistically equivalent graph representation in which it is a terminal node. In [9] the concept of a terminal node (sink) was applied to essential graphs. In Figure 1, for example, the nodes 1 and 3 are boundary variables. In an EP sequence boundary variables are very easy to identify.

Definition 6 (Boundary variable) *A variable that appears only one time in the whole sequence is a boundary variable.*

In a complete essential graph all vertices are boundary variables. Likewise, all vertices in an edgeless graph are boundary variables.

As it will be shown in the next section, boundary variables may be omitted in the process of model construction. After the deletion of boundary variables, however, new variables may appear that are in the boundary now because factors consisting of tails only should be removed. This leads us to the concept of *boundary layers* recursively defined as follows.

Definition 7 (Boundary layer) *Let S_1 be a perfect sequence. Its boundary layer 1 consists of the set of all boundary variables of S_1. S_i is obtained by deleting all boundary variables from S_{i-1}. Boundary layer i of the sequence S_1 consists of the set of all boundary variables of S_i.*

We order the variables of a perfect sequences such that variables which are members of a higher layer are before (left of) variables which are members of a lower layer. The variables are thus partially ordered by the boundary layers.

Example 3 The sequence $\{2\},\{3\},\{1,4\},\{1,3,4,5\},\{2,3,4,6\}$ has two boundary layers, $\{5,6\}$ and $\{1,2,3,4\}$.

4.1 Deletion of boundary variables

In this section we shall show one of the reasons why we are interested in boundary variables. For this we need the following two assertions proven in [4] and [5], respectively.

Theorem 3 *Let M be such that $K \cap L \subseteq M \subseteq L$; then*

$$\pi(x_K) \triangleright \kappa(x_L) = (\pi(x_K) \triangleright \kappa(x_M)) \triangleright \kappa(x_L).$$

Theorem 4 *If $K_1 \supseteq (K_2 \cap K_3)$ then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2.$$

If a variable X_ℓ is a *boundary variable* (in layer 1) than it need not to be taken into account in the process of model construction. If $\ell \in K_k$ and $\ell \notin K_1 \cup \ldots \cup K_{k-1} \cup K_{k+1} \cup \ldots \cup K_n$ then, due to Theorem 3,

$$\pi_1(x_{K_1}) \triangleright \ldots \triangleright \pi_k(x_{K_k}) = \pi_1(x_{K_1}) \triangleright \ldots \triangleright \pi_k(x_{K_k \setminus \{\ell\}}) \triangleright \pi_k(x_{K_k}),$$

and therefore also

$$\begin{aligned}\pi_1(x_{K_1}) \triangleright \ldots \triangleright \pi_k(x_{K_k}) \triangleright \pi_{k+1}(x_{K_{k+1}}) \\ = \pi_1(x_{K_1}) \triangleright \ldots \triangleright \pi_k(x_{K_k \setminus \{\ell\}}) \triangleright \pi_k(x_{K_k}) \triangleright \pi_{k+1}(x_{K_{k+1}}) \\ = \pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_k \triangleright \pi_{k+1}.\end{aligned}$$

Since $\ell \notin K_{k+1}$, it is obvious that

$$K_1 \cup \ldots \cup K_{k-1} \cup (K_k \setminus \{\ell\}) \supseteq (K_k \cap K_{k+1}),$$

and therefore applying Theorem 4 we get

$$\pi_1 \triangleright \ldots \triangleright \pi_{k+1} = \pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_{k+1} \triangleright \pi_k.$$

Similarly, $\ell \notin K_{k+2}$ and therefore

$$K_1 \cup \ldots \cup K_{k-1} \cup (K_k \setminus \{\ell\}) \cup K_{k+1} \supseteq (K_k \cap K_{k+2}),$$

and thus Theorem 4 yields

$$\pi_1 \triangleright \ldots \triangleright \pi_{k+2} = \pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_{k+1} \triangleright \pi_{k+2} \triangleright \pi_k.$$

In this way we can proceed further until we get

$$\pi_1 \triangleright \ldots \triangleright \pi_n = \pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_{k+1} \triangleright \ldots \triangleright \pi_n \triangleright \pi_k.$$

To summarize these steps: If $\ell \in K_k$ and $\ell \notin K_1 \cup \ldots \cup K_{k-1} \cup K_{k+1} \cup \ldots \cup K_n$ then $\pi_1 \triangleright \ldots \triangleright \pi_n$ is perfect if and only if $\pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_{k+1} \triangleright \ldots \triangleright \pi_n \triangleright \pi_k$ is perfect and both these sequences generate the same model. If the latter sequence is perfect then, obviously, $\pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_{k+1} \triangleright \ldots \triangleright \pi_n$ and π_k are consistent and therefore $\pi_1 \triangleright \ldots \triangleright \pi_{k-1} \triangleright \pi_k^{[\ell]} \triangleright \pi_{k+1} \triangleright \ldots \triangleright \pi_n \triangleright \pi_k$ is perfect, too. This means that if there are boundary variables in the system $\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})$ we can marginalize these variables out and neglect them during a process of a model construction.

5 Applications

5.1 Enumerating the model space

Comparing, evaluating, and learning models requires knowledge about the set of possible models. EP sequences and the boundary layers introduced in the previous sections facilitates the enumeration and counting methods of labeled and unlabeled model structures.

We treat the labeled structures first. For the ease of comparing sequences the following coding scheme is used. We encode each factor by the sum of powers of 2 of the labels (the natural numbers $1, \ldots, |N|$) of the variables it contains. Thus, we obtain $6, 8, 22$ for the sequence $\{1, 2\}, \{3\}, \{1, 2, 4\}$. This integer code allows to order sequences lexicographically. We read the sequence backwards, from right to left, and say that a sequence S_1 is lexicographically greater than S_2 if—when both sequences are rank ordered by the size of the integer codes of the factors—the greatest factor code that is different in both sequences is in S_1. An EP sequence is maximal if its integer code is lexicographically maximal. Ordering variables according to layers and maximal EP sequences introduces a (linear) *standard order* for variables in EP sequences and essential graphs. The standard order guarantees that the sequences (or the corresponding graphs) of equivalent models always are identical (or look the same). This helps to compare different networks.

With the help of the standard order we find the number of labeled models.

Theorem 5 (Number of labelings of a given EP sequence) *Let S be a EP sequence with N variables and r boundary layers and $\sigma_i,\ i = 1, \ldots, r$ variables in each layer. Let $\mathcal{P}$ be the set of $\sigma_1!\ \sigma_2\ ! \ldots \sigma_r!$ permutations generated by permuting the variables within each of the boundary layers. If $\mathcal{M}$ is the subset of $\mathcal{P}$ in which all its permutations generate the standard order, then the number of labellings is*

$$\frac{|N|!}{|\mathcal{M}|}$$

The theorem is an immediate consequence of the Burnside lemma [13].

Example 4 Figure 2 shows the twenty possible unlabeled essential graphs together with the number of possible labellings. The boundary variables belonging to the same layers are marked by the same symbols. For example, model (19) in Figure 2 has 6 labellings as the sequence $\{1\}, \{2\}, \{1, 2, 3, 4\}$ admits four permutations giving rise to the standard order with the integer code $2, 4, 30$. The permutations are $(1, 2, 3, 4)$, $(2, 1, 3, 4)$, $(1, 2, 4, 3)$, and $(2, 1, 4, 3)$. Because $4!/4 = 6$ we have 6 differently labeled models.

We now turn to the enumeration and number of *unlabeled* EP sequences. For this a method based on perfect sequences was described in [9]. The method counts the number of representatives (standard orders) for certain subsets of

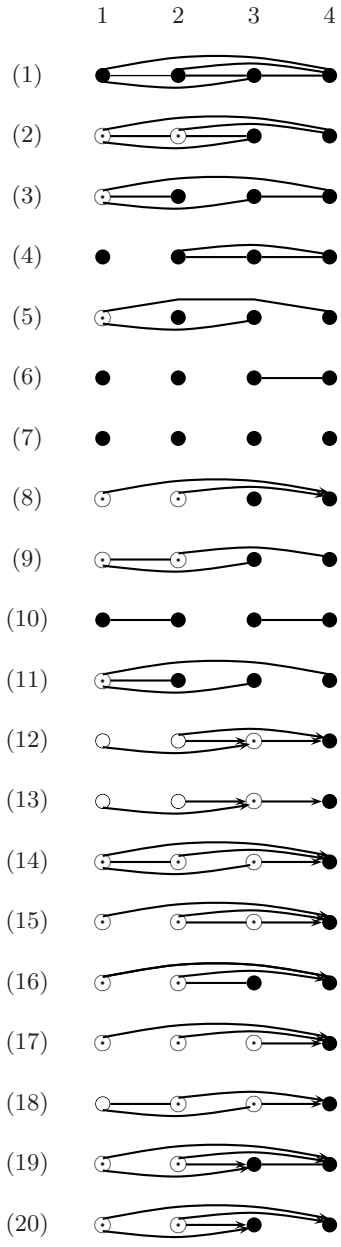

Figure 2: The twenty possible unlabeled essential graphs for $n = 4$, compare [1]. $\bullet, \odot$, and $\circ$ indicate vertices in boundary layer 1, 2, and 3, respectively. The first column at the right hand side of the graphs contains the number of possible labellings, the second column the perfect sequences corresponding to the graphs.

the model space. The subsets are generated by first selecting partitions of variables (corresponding to sequences containing heads only) and then by inserting systematically all possible tail variables.

Enumeration methods allow to establish statistics of many interesting features of models in the model space, such as the distribution component frequencies, the distribution of the number of edges etc.

5.2 Learning and evaluating models

Usually there is considerable prior knowledge about the CI model we want to learn from data. We expect that the variables are interrelated on an intermediate level, or that the number of factors is not extremely small or large. If we were, for example, convinced that all variables were independent, we would not analyze them together. Accordingly, the prior probability of such a model is small. The highly typical model structures might be evaluated differently when compared to rather infrequent and special structures.

Moreover, search procedures will profit from reasonable subdivisions of the model space. They are useful for traversing the model space. Prior probabilities can be an efficient guide in the search process.

Perfect sequences can be employed in several different way for traversing the model space for learning models from data. One method is to starts with searching for the best partition of the $|N|$ variables, i.e., factors with empty tails or, equivalently, completely connected components with undirected edges only. The process continues by inserting variables into tails and proceeds by adding more and more boundary layers. A different strategy starts from a cover of the set of all variable and proceeds by inserting and deleting factors and variables.

Learning requires local decisions about the dependence or independence of variables. Mutual information is a well founded measure of dependence. The statistical properties of mutual information were studied by Hutter [3], substantially improving an earlier proposal of the authors [7, 8]. They allow us to find the posterior probabilities that the mutual information in an individual factor in a perfect sequence is greater than a given threshold.

The *local* evaluation of factors, though, is not sufficient, a *global* criterion to evaluate the quality of a whole model is required. One of the best criteria for evaluating perfect sequence models is a Kullback-Leibler divergence of the model distributions from the unknown distribution, which generated data. It was already shown by Perez [12] that for a specific class of multidimensional probabilistic models this divergence minimizes for the distribution, which maximizes its informational content. This holds independently of the data generating distribution. This nice property holds also for perfect sequence models under the assumption that the low-dimensional distributions, from which the model is composed, are all marginals of the unknown generating distribution. Under this assumption, our goal is thus to maximize informational content of the model, which is defined for

$$\kappa(x_N) = \pi(x_{K_1}) \triangleright \ldots \triangleright \pi_n(x_{K_n})$$

by the expression

$$IC(\kappa(x_N)) = \sum_{x_N \in \mathbf{X}_N} \kappa(x_N) \log \frac{\kappa(x_N)}{\prod_{i \in N} \kappa(x_i)}.$$

If $\pi_1, \ldots, \pi_n$ is a perfect sequence, this value can easily be computed

$$IC(\kappa(x_N)) = \sum_{j=1}^{n} IC(\pi_j(x_{K_j})) \quad - \quad \sum_{j=2}^{n} IC(\pi_j(x_{T_j})),$$

where T_j denotes the tail of the factor K_j

$$T_j = K_j \cap (K_1 \cup \ldots \cup K_{j-1}).$$

Information content may be used as a global criterion to evaluate a whole perfect sequence model. Its statistical properties might be investigated on similar lines as Hutter [3] did with respect to mutual information.

Acknowledgements

The research was partially supported by the Austrian-Czech agency AKTION under the grant KONTAKT 2004/19, and AV ČR under grant A2075302.

References

[1] Andersson, S. A., Madigan, D., and Perlman, M. D. (1998). A characterization of Markov equivalence classes for acyclic digraphs. Annals of Statistics, **25**, 505-541.

[2] 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. Stat.* **11** (1940), pp. 427-444.

[3] Hutter, M. (2001). Distribution of mutual information. IDSIA, CH-6928 Galleria 2, Manno-Lugano, Switzerland, `marcus@idsia.ch`

[4] Jiroušek, R. (1997), Composition of probability measures on finite spaces. In: Geiger, D., Shenoy, P.P. (eds.) *Uncertainty in Artificial Intelligence.* San Francisco, Morgan Kaufman Publ., pp 274-281.

[5] Jiroušek, R. (2000), Marginalization in composed probabilistic models. In: Boutilier, Craig, Goldszmidt, Moiss (ed.). *Uncertainty in artificial intelligence.* San Francisco, Morgan Kaufmann Publ., pp 301-308.

[6] Jiroušek, R. (2004), What is the difference between Bayesian networks and compositional models? In: H. Noguchi, H. Ishii, M. Inuiguchi (eds.) *Proceedings of the 7th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty.* Awaji Yumebutai ICC, Japan, pp 191-196.

[7] Kleiter, G. D. and Jiroušek, R. (1996). Learning Bayesian networks under the control of mutual information. In Moral, S. et al. (eds.). Sixth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, 985-990.

[8] Kleiter, G. D. (1999). The posterior probability of Bayes nets with strong dependences. *Soft Computing*, **3**, 162-173.

[9] Kleiter, G. D. (2004). Enumerating equivalence classes of Bayesian networks by perfect sequences. In Bouchon-Meunier, B., Coletti, G. and Yager, R. R. (eds.), *Proceedings of the Tenth Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Vol. 2*, Editrice Universita La Sapienza, Italy, pp 1287–1294.

[10] Malvestuto F.M. (1988), Existence of extensions and product extensions for discrete probability distributions, *Discrete Mathematics* 69, pp 61-77.

[11] Malvestuto F.M. (1989), Computing the maximum-entropy extension of given discrete probability distributions, *Computational Statistics & Data Analysis* 13, pp 299-311.

[12] Perez, A. (1977). ε-admissible simplication of the dependence structure of a set of random variables. *Kybernetika*, **13**, pp. 439-450.

[13] Williamson, S. G. (1985), Combinatorics for Computer Science. Rockville, ML, Computer Science Press.