M. Studený Pojem multiinformace v pravděpodobnostním rozhodování (in Czech, translation: The notion of multiinformation in probabilistic decision making), doctoral thesis, Czechoslovak Academy of Sciences, Institute of Information Theory and Automation, Prague 1987, Czechoslovakia, 121 pages.

ABSTRACT Let $\{(\mathbf{X}_i, \mathcal{X}_i)\,; i \in N\}$ be a collection of measurable spaces, indexed by a finite non-empty set N, $\mathcal{A}$ a system of non-empty subsets of N, and $\mathcal{P} = \{P_A\,; A \in \mathcal{A}\}$ a collection of probability measures where P_A is defined on the measurable space $(\mathbf{X}_A, \mathcal{X}_A)$ (this denotes the product of the measurable spaces $(\mathbf{X}_i, \mathcal{X}_i)$ for $i \in A$).

The first part of the thesis concerns the question of the existence of a (joint) probability distribution P on $(\mathbf{X}_N, \mathcal{X}_N)$ having measures from $\mathcal{P}$ as marginal measures (that is, the *consistency* of $\mathcal{P}$) and the question how to verify the consistency. This problem is sometimes called the *marginal problem.* The first chapter is a survey of the results of a research report without proofs. The condition of *consonancy*, that is, for every couple $P_A, P_B \in \mathcal{P}$ the marginal measures of P_A and P_B on $(\mathbf{X}_{A\cap B}, \mathcal{X}_{A\cap B})$ coincide, is a clear necessary but not a sufficient condition for consistency. A necessary and sufficient condition valid under certain topological assumptions on $\mathbf{X}_i$, $i \in N$ is formulated. Because this condition is difficult to verify (it concerns a quite wide class of continous bounded functions on $\mathbf{X}_A$, $A \in \mathcal{A}$) attention is devoted to the question when $\mathcal{A}$ is *solvable*, that is, for every system $\mathcal{P} = \{P_A\,; A \in \mathcal{A}\}$ consonancy already implies consistency. Under a certain topological assumptions on $\mathbf{X}_i$, $i \in N$, it is shown that $\mathcal{A}$ is solvable iff it satisfies the *running intersection property*, that is, can be ordered into a sequence $A_1, \ldots, A_n$, $n \geq 1$ such that $\forall i > 1\ \exists j < i\ \ A_i \cap (\bigcup_{k<i} A_k) \subset A_j$. An example that if the topological assumptions are omitted, then the conclusion may not not hold, is given.

The second chapter is devoted to a few concepts from probability theory. By a *channel* (= Markov kernel) from a measurable space $(\mathbf{X}, \mathcal{X})$ to a measurable space $(\mathbf{Y}, \mathcal{Y})$ is understood a collection of probability measures $\{P_x\,; x \in X\}$ on $(\mathbf{Y}, \mathcal{Y})$ such that for every fixed $A \in \mathcal{Y}$, $P_x(A)$ as a function of x is $\mathcal{X}$-measurable. It is *dominated* if there exists a measure μ on $(\mathbf{Y}, \mathcal{Y})$ such that P_x is absolutely continuous with respect to μ for every $x \in X$. Distributions created by a channel are introduced. A probability measure P on $(\mathbf{X} \times \mathbf{Y}, \mathcal{X} \times \mathcal{Y})$ is *marginally continuous* if it is absolutely continous with respect to the product measure $P^{\mathbf{X}} \times P^{\mathbf{Y}}$ where $P^{\mathbf{X}}$, $P^{\mathbf{Y}}$ are respective marginals of P on $(\mathbf{X}, \mathcal{X})$ and $(\mathbf{Y}, \mathcal{Y})$. Two equivalent characterization of marginally continous measures are derived:

- P is absolutely continous with respect to the product measure $\mu \times \nu$ where μ is an arbitrary mearure on $(\mathbf{X}, \mathcal{X})$ and ν an arbitrary measure on $(\mathbf{Y}, \mathcal{Y})$,
- P is created by a dominant channel from $(\mathbf{X}, \mathcal{X})$ to $(\mathbf{Y}, \mathcal{Y})$.

The rest of the section is devoted to the concept of (regular) conditional probability and to the concept of *conditional product measure.*

The third chapter deals with concepts from information theory. First, basic information theoretical concepts of relative entropy and conditional mutual information are recalled. Given a simultaneous probability measure P on $(\mathbf{X}_N, \mathcal{X}_N)$ the *multiinformation function* induced by P is a non-negative real function on the power set of N which ascribes to non-empty $A \subset N$ the multiinformation of the marginal P^A on $(\mathbf{X}_A, \mathcal{X}_A)$, that is the relative entropy of P^A with respect to the product measure $\prod_{i\in A} P^i$ of one-dimensional marginals

P^i of P^A on $(\mathbf{X}_i, \mathcal{X}_i)$, $i \in A$. A formula for the conditional mutual information on basis of the multiinformation function is derived. It is shown that the conditional product of two consonant measures P_A, P_B on $(\mathbf{X}_A, \mathcal{X}_A)$ and $(\mathbf{X}_B, \mathcal{X}_B)$ where $A \cup B = N$ solves the problem of minimization of the value of multiinformation of P on $(\mathbf{X}_N, \mathcal{X}_N)$ within the class of distributions P having P^A and P^B as marginals.
In case that a collection of probability measures $\mathcal{P} = \{P_A\,; A \in \mathcal{A}\}$ is consistent the following question arises in connection with practical applications: how to choose and how to store in memory of a computer the corresponding joint measure P? An approach based on the concept of *dependence structure simplification* introduced by A. Perez is presented in the fouth chapter. The original definition is modified and refined. Informally, for every ordering $A_1, \ldots, A_n$, $n \geq 1$ of sets in $\mathcal{A}$ the formula $\bar{P} = \prod_{i=1}^{n} P_{A_i|\cup_{k<i} A_k}$ where $P_{A_i|\cup_{k<i} A_k}$ is a channel which creates P_{A_i} (= regular conditional probability computed from P_{A_i}), defines a certain probability measure $\bar{P}$ which is supposed to be an approximation of an unknown joint measure P having $\mathcal{P}$ as marginals. A formula expressing the relative entropy $H(P, \bar{P})$ of P with respect to $\bar{P}$ is derived under assumptions involving marginal continuity of measures in $\mathcal{P}$. This simplifies the question of finding the *optimal* dependence structure simplification, that is $\bar{P}$ minimizing $H(P, \bar{P})$ within the class of $\bar{P}$'s. It is shown that the choice of this optimal approximation does not depend on the choice of a simultaneous measure P, but on $\mathcal{P}$ only: the optimal approximations are characterized by those orderings of $\mathcal{A}$ which minimize certain functional defined by means of the multiinformation functions for P_A, $A \in \mathcal{A}$. The last result of the this chapter says that if there exists a dependence structure simplification $\bar{P}$ having $\mathcal{P}$ as marginals, then it is uniquely determined and optimal in the above described sense.
The last chapter is devoted to the methods of finding the optimal approximation. An algorithm, based on the application of the multiinformation function, which finds the optimal ordering of $\mathcal{A}$ under assumption that every set of $\mathcal{A}$ intersects at most three other sets of $\mathcal{A}$, is presented. The algorithm is recursive and based on certain simple combinatorial results concerning hypergraphs (that is, systems $\mathcal{A}$ of subsets of a finite non-empty set N). The chapter contains a couple of counterexamples illustrating technical remarks related to the algorithm.