Divergence from factorizable distributions and matroid representations by partitions

František Matúš

***Abstract*— Maximization of the information divergence from any hierarchical log-linear model is studied. A new upper bound on the maximum is presented and its tightness analyzed. For the models given by the bases of a matroid, the latter is related to matroid representations by partitions or, equivalently, to ideal secret sharing schemes. A new link between the divergence maximization, the maximum likelihood principle and secret sharing is established.**

***Index Terms*— Information divergence, relative entropy, Shannon entropy, exponential family, hierarchical model, log-linear model, contingency table, Gibbs distribution, matroid representation, partition, secret sharing scheme, almost affine code, maximum likelihood.**

I. Introduction

Let N be a finite set and $\mathcal{A}$ a nonempty family of subsets of N. A probability measure (pm) P on the Cartesian product $X = \prod_{i\in N} X_i$ of finite state spaces X_i is called $\mathcal{A}$-*factorizable* if there exist real functions ψ_I on $X_I = \prod_{i\in I} X_i$ such that

$$P(x) = \prod_{I\in\mathcal{A}} \psi_I(\pi_I x)\,, \qquad x \in X\,,$$

where π_I projects x to X_I. The family of all $\mathcal{A}$-factorizable pm's on X that are positive, in the sense $P(x) > 0$ for all $x \in X$, is denoted here by $\mathcal{E}_{N,\mathcal{A},X}$. Statisticians speak about a hierarchical log-linear model [10] as the family does not change if $\mathcal{A}$ is enriched by subsets of the sets from $\mathcal{A}$.

The *information divergence*, or relative entropy, between pm's P, Q on X is given by

$$D(P\|Q) = \begin{cases} \sum_{x\in s(P)} P(x)\,\ln\frac{P(x)}{Q(x)}\,, & s(P)\subseteq s(Q)\,,\\ +\infty\,, & \text{otherwise,}\end{cases}$$

where $s(P) = \{x \in X : P(x) > 0\}$ is the support of P, and the divergence of P from the family $\mathcal{E} = \mathcal{E}_{N,\mathcal{A},X}$ by

$$D(P\|\mathcal{E}) = \inf_{Q\in\mathcal{E}} D(P\|Q)\,.$$

This work focuses on maximization of the function

$$D(\cdot\|\mathcal{E})\colon P \mapsto D(P\|\mathcal{E})$$

over P in the simplex of all pm's on X. If $\mathcal{E}$ denotes instead an arbitrary exponential family of pm's on a finite set this maximization problem goes back to [1], for later progress see [14], [15], [2], [17].

Manuscript submitted November 2, 2008; revised June 3, 2009. This work was partially supported by Grant Agency of Academy of Sciences of the Czech Republic under the grant IAA 100750603, and by Grant Agency of the Czech Republic under the grant 201/04/0393. F. Matúš is with Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic (e-mail: matus@utia.cas.cz).

Theorem 1 of Section III presents upper bounds on the divergence $D(P\|\mathcal{E}_{N,\mathcal{A},X})$ by sums of Shannon entropies of marginals of P and the cardinalities of state spaces. As a consequence, a new upper bound on $D(\cdot\|\mathcal{E}_{N,\mathcal{A},X})$ emerges. Necessary conditions for tightness of the bound are formulated in Theorem 2.

In Section IV, the attainment of the upper bound is discussed when $\mathcal{A}$ is the family of bases of a matroid with the ground set N and the state spaces have the same cardinality d. When the matroid has no coloop and the bound is tight, maximizers of $D(\cdot\|\mathcal{E}_{N,\mathcal{A},X})$ have a special form. In particular, they can be constructed from matroid representations by partitions of the degree d or, equivalently, by ideal secret sharing schemes with the secret of the size d.

Statistical and cryptographic interpretations of these results are collected in Section V, viewing the maximizers as empirical measures and applying the maximum likelihood principle.

II. Preliminaries

In the sequel, elements and singletons of N are frequently not distinguished. For example, $\{i\} \cup I$ shortens to $i \cup I$ and $\pi_{\{i\}}$ to π_i, $i\in N$, $I\subseteq N$. The union $\bigcup_{I\in\mathcal{A}} I$ is denoted by $\bigcup\mathcal{A}$.

Extending elementary notions from the matroid theory, an element of N not in $\bigcup\mathcal{A}$ is called a *loop* of $(N,\mathcal{A})$ and an element which belongs to every inclusion maximal set in $\mathcal{A}$ is called a *coloop* of $\mathcal{A}$.

In the case $I = \emptyset$, the space $X_\emptyset$ is assumed to be a singleton, say $X_\emptyset = \{\emptyset\}$. Thus, $\pi_\emptyset$ projects $x \in X$ to $\emptyset$ and $X_\emptyset$ supports a unique pm.

A. Linear and exponential families

For $I \subseteq N$ and $x_I, y_I \in X_I$ let δ_{x_I,y_I} equal one if $x_I = y_I$ and zero otherwise. The *marginal* of a pm P on X to I is the pm $\pi_I P$ on X_I given by

$$\pi_I P(x_I) = \sum_{y\in X} P(y)\,\delta_{x_I,\pi_I y}\,, \qquad x_I \in X_I\,.$$

The linear family $\mathcal{L}_{P,\mathcal{A}}$ determined by a pm P and $\mathcal{A}$ is defined as the set of all pm's Q on X with $\pi_I Q = \pi_I P$ for $I \in \mathcal{A}$.

As well known, $\mathcal{E}_{N,\mathcal{A}} = \mathcal{E}_{N,\mathcal{A},X}$ is the full exponential family determined by the counting measure on X and a canonical statistic f that takes values in the Euclidean space $\mathbb{R}^{X_\mathcal{A}}$ where $X_\mathcal{A} = \{(I, y_I)\colon I \in \mathcal{A}\,,\ y_I \in X_I\}$. The statistic can be given by

$$f(x) = \big(\delta_{y_I,\pi_I x}\colon (I, y_I) \in X_\mathcal{A}\big)\,, \quad x \in X\,.$$

Thus, the pm's in $\mathcal{E}_{N,\mathcal{A}}$ have the exponential representation

$$P(x) = C_P \cdot \exp\Big[\sum_{(I,y_I)\in X_\mathcal{A}} \vartheta_{I,y_I}\delta_{y_I,\pi_I x}\Big]\,, \quad x \in X\,,$$

where $\vartheta = (\vartheta_{I,y_I})_{(I,y_I)\in X_\mathcal{A}} \in \mathbb{R}^{X_\mathcal{A}}$ is an arbitrary vector parameter and C_P a corresponding normalizing constant. The bracket equals the scalar product of ϑ and $f(x)$ in $\mathbb{R}^{X_\mathcal{A}}$ and simplifies to $\sum_{I\in\mathcal{A}} \vartheta_{I,\pi_I x}$. Under the formal substitution $\vartheta_{I,y_I} = \ln\psi_I(y_I)$, exponential representations of positive pm's P rewrite to multiplicative factorizations and vice versa. Standard references to the exponential families include [3], [4], [5]. For the approach through toric algebra see [9].

The closure of the family $\mathcal{E}_{N,\mathcal{A}}$ is denoted by $\mathit{cl}(\mathcal{E}_{N,\mathcal{A}})$. The following assertions are well-known and their elementary proofs can be found in [8] or [11, Proposition 4]. For general results in this direction see [6].

Lemma 1: $D(\cdot\|\mathcal{E}_{N,\mathcal{A}}) = D(\cdot\|\mathit{cl}(\mathcal{E}_{N,\mathcal{A}}))$

Lemma 2: If P is any pm on X then $\mathcal{L}_{P,\mathcal{A}}$ and $\mathit{cl}(\mathcal{E}_{N,\mathcal{A}})$ have a single pm Q in common, $D(P\|Q) = D(P\|\mathcal{E}_{N,\mathcal{A}})$ and Q is unique in $\mathit{cl}(\mathcal{E}_{N,\mathcal{A}})$ with this property.

B. First decomposition lemma

The following assertion is a key component of the proof of Theorem 1. Let $\mathit{re}_L\,\mathcal{A} = \{L\cap K\colon K\in\mathcal{A}\}$ denote the restriction of $\mathcal{A}$ onto $L\subseteq N$.

Lemma 3: If $N = I\cup J$ and a set in $\mathcal{A}$ covers $I\cap J$ then for any pm P on X

$$\begin{aligned} D(P\|\mathcal{E}_{N,\mathcal{A}}) \leqslant{}& D(\pi_I P\|\mathcal{E}_{I,\mathit{re}_I\,\mathcal{A}}) + D(\pi_J P\|\mathcal{E}_{J,\mathit{re}_J\,\mathcal{A}}) \\ &+H(\pi_I P)+H(\pi_J P)-H(P)-H(\pi_{I\cap J}P)\,. \end{aligned} \tag{1}$$

Additionally, if each set from $\mathcal{A}$ is contained in I or J then (1) becomes tight.

Proof: In this proof $K = I$ or $K = J$. By Lemma 2,

$$D(\pi_K P\|Q_K) = D(\pi_K P\|\mathcal{E}_{K,\mathit{re}_K\,\mathcal{A}}) \tag{2}$$

where the pm Q_K satisfies

$$Q_K \in \mathcal{L}_{\pi_K P,\mathit{re}_K\,\mathcal{A}} \tag{3}$$
$$Q_K \in \mathit{cl}(\mathcal{E}_{K,\mathit{re}_K\,\mathcal{A}})\,. \tag{4}$$

Since $I\cap J$ is contained in some $L\in\mathcal{A}$ it is also contained in $L\cap K\in \mathit{re}_K\,\mathcal{A}$. This and (3) imply that the marginals $\pi_{I\cap J}P$, $\pi_{I\cap J}Q_I$ and $\pi_{I\cap J}Q_J$ coincide.

Then, with the notation

$$R(x) = \frac{Q_I(\pi_I x)\cdot Q_J(\pi_J x)}{\pi_{I\cap J}P(\pi_{I\cap J}x)}$$

if x projects to $\mathit{s}(\pi_{I\cap J}P)$ and $R(x)=0$ otherwise, R is a pm on X. It is not difficult to show that $R\in\mathit{cl}(\mathcal{E}_{N,\mathcal{B}})$ where $\mathcal{B} = \mathit{re}_I\,\mathcal{A}\cup \mathit{re}_J\,\mathcal{A}$. In fact, (4) implies that Q_K is the limit of a sequence $Q_{K,n}$ of positive $(\mathit{re}_K\,\mathcal{A})$-factorizable pm's whence the positive pm's given by

$$R_n(x) = \frac{Q_{I,n}(\pi_I x)\cdot Q_{J,n}(\pi_J x)}{\pi_{I\cap J}Q_{J,n}(\pi_{I\cap J}x)}\,, \qquad x\in X\,,$$

are $\mathcal{B}$-factorizable. Since $R_n(x)\to R(x)$ when x projects to $\mathit{s}(\pi_{I\cap J}P)$ the pm's R_n converge to R.

Any set of $\mathcal{B}$ is covered by a set of $\mathcal{A}$. Therefore, $\mathcal{E}_{N,\mathcal{B}}$ is contained in $\mathcal{E}_{N,\mathcal{A}}$. Lemma 1 and this containment imply that $D(P\|\mathcal{E}_{N,\mathcal{A}})$ is equal to $D(P\|\mathit{cl}(\mathcal{E}_{N,\mathcal{A}}))$ which is upper bounded by $D(P\|\mathit{cl}(\mathcal{E}_{N,\mathcal{B}}))$. This is majorized by $D(P\|R)$ because $R\in\mathit{cl}(\mathcal{E}_{N,\mathcal{B}})$. Hence, it suffices to show that $D(P\|R)$ equals the right-hand side of ineq. (1).

The support of R contains $\mathit{s}(P)$ because $\mathit{s}(\pi_K P)$ is a subset of $\mathit{s}(Q_K)$ by finiteness in (2). Thus, $D(P\|R)$ rewrites to

$$\sum_{x\in\mathit{s}(P)} P(x)\ln\frac{\pi_I P(\pi_I x)\cdot\pi_J P(\pi_J x)\cdot P(x)\cdot\pi_{I\cap J}P(\pi_{I\cap J}x)}{Q_I(\pi_I x)\cdot Q_J(\pi_J x)\cdot\pi_I P(\pi_I x)\cdot\pi_J P(\pi_J x)}\,.$$

Then, ineq. (1) follows on account of (2).

If each set of $\mathcal{A}$ is covered by I or J then $\mathcal{E}_{N,\mathcal{A}} = \mathcal{E}_{N,\mathcal{B}}$. Thus $R\in\mathit{cl}(\mathcal{E}_{N,\mathcal{A}})$. The incidence $R\in\mathcal{L}_{P,\mathcal{A}}$ follows by combining (3), $\pi_I R = Q_I$ and $\pi_J R = Q_J$. By Lemma 2, $D(P\|R)$ equals $D(P\|\mathcal{E}_{N,\mathcal{A}})$ whence (1) becomes tight. ∎

Corollary 1: If $\bigcup\mathcal{A}\subseteq K\subseteq N$ then for any pm P on X

$$\begin{aligned} D(P\|\mathcal{E}_{N,\mathcal{A}}) ={}& D(\pi_K P\|\mathcal{E}_{K,\mathcal{A}}) \\ &+H(\pi_K P)-H(P)+\ln|X_{N\setminus K}|\,. \end{aligned} \tag{5}$$

Proof: Lemma 3 is applied to $I = N\setminus K$ and $J = K$, and provides (1) which is tight by the assumption on K. Since $\mathit{re}_I\,\mathcal{A} = \{\emptyset\}$ it follows that $\mathcal{E}_{I,\mathit{re}_I\,\mathcal{A}}$ consists of the uniform pm R_I on X_I and

$$D(\pi_I P\|\mathcal{E}_{I,\mathit{re}_I\,\mathcal{A}}) = D(\pi_I P\|R_I) = -H(\pi_I P)+\ln|X_I|\,.$$

This and the tight ineq. (1) sum to (5). ∎

Remark 1: By [7, Section 5], $D(\cdot\|\mathcal{E}_{N,\mathcal{A}})$ is continuous on the simplex of all pm's on X. This and Corollary 1 imply

$$\max D(\cdot\|\mathcal{E}_{N,\mathcal{A}}) = \max D(\cdot\|\mathcal{E}_{K,\mathcal{A}}) + \ln|X_{N\setminus K}| \tag{6}$$

assuming $K = \bigcup\mathcal{A}$. In fact, to prove the inequality $\leqslant$ in (6), sum $H(\pi_K P)\leqslant H(P)$ with eq. (5) and maximize. To prove the opposite one, any maximizer Q_K of $D(\cdot\|\mathcal{E}_{K,\mathcal{A}})$ is extended to a pm P such that $H(Q_K) = H(P)$ and then (5) is applied.

Combining (5) and (6), a pm P maximizes $D(\cdot\|\mathcal{E}_{N,\mathcal{A}})$ if and only if the marginal $\pi_K P$ maximizes $D(\cdot\|\mathcal{E}_{K,\mathcal{A}})$ and $H(\pi_K P)$ equals $H(P)$. Roughly speaking, the loops can be removed from the family $\mathcal{A}$ in the maximization problem.

C. Second decomposition lemma

For a pm P on X, $I\subseteq N$, $J = N\setminus I$ and $x_I\in\mathit{s}(\pi_I P)$ let $P(\cdot|x_I)$ denote the pm on X_J that assigns to $x_J\in X_J$ the ratio $P(x)/\pi_I P(x_I)$ where $x\in X$ is given by the projections $\pi_I x = x_I$ and $\pi_J x = x_J$.

Lemma 4: If I is a set of coloops of $\mathcal{A}$ and $J = N\setminus I$ then for any pm P

$$D(P\|\mathcal{E}_{N,\mathcal{A}}) = \textstyle\sum_{x_I\in\mathit{s}(\pi_I P)} D(P(\cdot|x_I)\|\mathcal{E}_{J,\mathit{re}_J\,\mathcal{A}})\,\pi_I P(x_I)\,.$$

Proof: By Lemma 2, for $x_I\in\mathit{s}(\pi_I P)$ a unique pm

$$Q_J^{x_I} \in \mathcal{L}_{P(\cdot|x_I),\mathit{re}_J\,\mathcal{A}}\cap\mathit{cl}(\mathcal{E}_{J,\mathit{re}_J\,\mathcal{A}}) \tag{7}$$

exists such that

$$D(P(\cdot|x_I)\|Q_J^{x_I}) = D(P(\cdot|x_I)\|\mathcal{E}_{J,\mathit{re}_J\,\mathcal{A}})\,. \tag{8}$$

Let $R(x) = \pi_I P(\pi_I x)\cdot Q_J^{\pi_I x}(\pi_J x)$ if x projects to $\mathit{s}(\pi_I P)$ and $R(x) = 0$ otherwise. Then, R is a pm on X with $\mathit{s}(P)\subseteq\mathit{s}(R)$ using that $\mathit{s}(P(\cdot|x_I))\subseteq\mathit{s}(Q_J^{x_I})$ on account of finiteness in (8).

Obviously, $\pi_I R = \pi_I P$ and $R(\cdot|x_I) = Q_J^{x_I}$ for $x_I \in s(\pi_I P)$. These equalities, (8) and the simple identity

$$D(P\|R) = D(\pi_I P\|\pi_I R) + \sum_{x_I \in s(\pi_I P)} D(P(\cdot|x_I)\|R(\cdot|x_I))\, \pi_I P(x_I)$$

imply

$$D(P\|R) = \sum_{x_I \in s(\pi_I P)} D(P(\cdot|x_I)\|\mathcal{E}_{J,re_J\,\mathcal{A}})\, \pi_I P(x_I)\,.$$

Due to Lemma 2, it remains to show that $R \in \mathcal{L}_{P,\mathcal{A}} \cap cl(\mathcal{E}_{N,\mathcal{A}})$.

To prove $R \in \mathcal{L}_{P,\mathcal{A}}$, it suffices to verify that $\pi_K R = \pi_K P$ for every inclusion maximal set K in $\mathcal{A}$. Since I is a set of coloops $I \subseteq K$ and for $x_K \in X_K$

$$\pi_K R(x_K) = \pi_I P(\pi_I x_K) \cdot \pi_{K\cap J} Q_J^{\pi_I x_K}(\pi_{K\cap J} x_K)$$

whenever x_K projects to $s(\pi_I P)$ and $\pi_K R(x_K) = 0$ otherwise. In the former case, $\pi_{K\cap J} Q_J^{\pi_I x_K}$ equals $\pi_{K\cap J} P(\cdot|\pi_I x_K)$ by (7), and then $\pi_K R(x_K) = \pi_K P(x_K)$. This equality holds trivially in the latter case.

By (7), if $x_I \in s(\pi_I P)$ then $Q_J^{x_I}$ is the limit of a sequence of positive $(re_J\,\mathcal{A})$-factorizable pm's $Q_{J,n}^{x_I}$. Otherwise, let $Q_{J,n}^{x_I}$ be uniform on X_J. Then, the positive pm's given by

$$R_n(x) = \left[(1-\tfrac{1}{n})\pi_I P(\pi_I x) + \tfrac{1}{n}|X_I|^{-1}\right] Q_{J,n}^{\pi_I x}(\pi_J x)\,,$$

for $x \in X$, are $\mathcal{A}$-factorizable and converge to R. ∎

Remark 2: It is a simple consequence of Lemma 4 that

$$\max D(\cdot\|\mathcal{E}_{N,\mathcal{A}}) = \max D(\cdot\|\mathcal{E}_{J,re_J\,\mathcal{A}})\,. \tag{9}$$

Combining the equality of Lemma 4 and (9), a pm P maximizes $D(\cdot\|\mathcal{E}_{N,\mathcal{A}})$ if and only if $P(\cdot|x_I)$ maximizes $D(\cdot\|\mathcal{E}_{J,re_J\,\mathcal{A}})$ for each x_I in the support of $\pi_I P$. Roughly speaking, the coloops can be removed from the family $\mathcal{A}$ in the maximization problem.

III. The upper bounds on $D(\cdot\|\mathcal{E}_{N,\mathcal{A}})$

The following result is based on a repeated use of Lemma 3.

Theorem 1: Let N be a finite set and $\mathcal{A}$ a nonempty family of its subsets. If $t \geqslant 1$, a set I_s with any $1 \leqslant s \leqslant t$ is covered by a set of $\mathcal{A}$, and the union of these sets equals $\bigcup\mathcal{A}$ then for any pm P on X

$$D(P\|\mathcal{E}_{N,\mathcal{A}}) \leqslant \ln|X_{N\setminus\bigcup\mathcal{A}}| + \sum_{s=2}^{t} H(\pi_{I_s\setminus I^s} P) \tag{10}$$

where $I^s = I_1 \cup \ldots \cup I_{s-1}$. If the equality takes place here then $H(P) = H(\pi_{I_1} P)$, and for $2 \leqslant s \leqslant t$

$$H(\pi_{I_s} P) = H(\pi_{I_s\cap I^s} P) + H(\pi_{I_s\setminus I^s} P)\,. \tag{11}$$

Proof: For any $1 \leqslant s \leqslant t$ the function $D(\cdot\|\mathcal{E}_{I_s,re_{I_s}\mathcal{A}})$ is identically zero because I_s is covered by a set L_s of $\mathcal{A}$. This and Lemma 3, with $I^{s+1} = I^s \cup I_s$ in the role of $N = I \cup J$ and $re_{I^{s+1}}\,\mathcal{A}$, containing the set $I^{s+1} \cap L_s$ that covers $I^s \cap I_s$, imply

$$D(\pi_{I^{s+1}} P\|\mathcal{E}_{I^{s+1},re_{I^{s+1}}\mathcal{A}}) \leqslant D(\pi_{I^s} P\|\mathcal{E}_{I^s,re_{I^s}\mathcal{A}}) + H(\pi_{I^s} P) + H(\pi_{I_s} P) - H(\pi_{I^{s+1}} P) - H(\pi_{I^s\cap I_s} P)\,.$$

By the basic inequalities for Shannon entropies [20], the second line is upper bounded by $H(\pi_{I_s\setminus I^s} P)$, tightly if and only if eq. (11) holds and

$$H(\pi_{I^s} P) = H(\pi_{I^{s+1}} P)\,. \tag{12}$$

Denoting $I^{t+1} = \bigcup\mathcal{A}$ by K, if follows that

$$D(\pi_K P\|\mathcal{E}_{K,re_K\,\mathcal{A}}) \leqslant \sum_{s=2}^{t} H(\pi_{I_s\setminus I^s} P)\,.$$

This, Corollary 1 and $H(\pi_K P) \leqslant H(P)$ imply ineq. (10). In the case of equality, eq. (11) and (12) are true for $2 \leqslant s \leqslant t$, and $H(\pi_K P) = H(P)$. Therefore, $H(P) = H(\pi_{I_1} P)$. ∎

Corollary 2: For every $I \in \mathcal{A}$

$$D(P\|\mathcal{E}_{N,\mathcal{A}}) \leqslant \ln|X_{N\setminus\bigcup\mathcal{A}}| + \sum_{k\in\bigcup\mathcal{A}\setminus I} H(\pi_k P)\,. \tag{13}$$

Corollary 3: $D(\cdot\|\mathcal{E}_{N,\mathcal{A}}) \leqslant \ln \min_{I\in\mathcal{A}} |X_{N\setminus I}|$

Remark 3: For the family $\mathcal{A}$ consisting of all singletons $\{i\}$, $i \in N$, the divergence of a pm P from the family $\mathcal{E}_{N,\mathcal{A}}$ equals $\sum_{i\in N} H(\pi_i P) - H(P)$, the multiinformation of P. Corollary 3 appeared as [2, Lemma 4.1] while ineq. (13) was contained in the proof of this lemma.

The remaining part of this section discusses tightness of the bound of Corollary 3.

Lemma 5: If a pm P has the marginal $\pi_J P$ uniform on X_J for every $J \in \mathcal{A}$, and $H(\pi_I P) = H(P)$ for some $I \in \mathcal{A}$ then

$$D(P\|\mathcal{E}_{N,\mathcal{A}}) = \ln \min_{I\in\mathcal{A}} |X_{N\setminus I}|\,. \tag{14}$$

Proof: Lemma 2 and the first assumption imply that $D(P\|\mathcal{E}_{N,\mathcal{A}})$ equals $D(P\|R)$ where R denotes the uniform pm on X. The latter divergence equals $\ln|X_N| - H(P)$. By the second assumption, $H(P)$ equals $H(\pi_I P)$ for some $I \in \mathcal{A}$. Hence $D(P\|\mathcal{E}_{N,\mathcal{A}}) = \ln|X_{N\setminus I}|$, and the assertion follows by Corollary 3. ∎

The above sufficient condition for a pm P to satisfy (14) is not necessary in general. In the special case of the family $\mathcal{A}$ considered in Remark 3, a different necessary and sufficient condition is described in [2, Theorem 4.3 and 4.4]. That condition, however, reduces to the above one when the state spaces have the same cardinality by [2, Corollary 4.10]. Under this assumption, the assertion of Lemma 5 can be reversed as follows.

Theorem 2: Let $t = \max_{I\in\mathcal{A}} |I|$, $\mathcal{A}^* = \{I \in \mathcal{A}\colon |I| = t\}$ and the state spaces X_i, $i \in N$, have the same cardinality d. If $D(P\|\mathcal{E}_{N,\mathcal{A}}) = [\,|N| - t\,]\ln d$ for a pm P on X then $H(P)$ equals $H(\pi_I P)$ for all $I \in \mathcal{A}^*$. If additionally $\mathcal{A}^*$ has no coloop then $\pi_J P$ is uniform for all $J \in \mathcal{A}$.

Proof: Assuming a sequence of sets I_s with $I_1 \in \mathcal{A}^*$ as in Theorem 1, the assumptions on $|X_i|$, $i \in N \setminus \bigcup\mathcal{A}$, and on P imply that ineq. (10) rewrites to

$$[\,|\bigcup\mathcal{A}| - t\,]\ln d \leqslant \sum_{s=2}^{t} H(\pi_{I_s\setminus I^s} P)\,.$$

Since the above differences $I_s \setminus I^s$ are pairwise disjoint, cover $\bigcup\mathcal{A} \setminus I_1$, $|I_1| = t$ and $|X_i| = d$ for $i \in \bigcup\mathcal{A}$ this inequality is tight and each marginal $\pi_{I_s\setminus I^s} P$ is uniform. Then, ineq. (10) is tight and it follows from the second assertion of Theorem 1 that $H(P) = H(\pi_{I_1} P)$ and

$$H(\pi_{I_2} P) = H(\pi_{I_2\cap I_1} P) + H(\pi_{I_2\setminus I_1} P)\,, \tag{15}$$

provided $t \geqslant 2$.

For any $I \in \mathcal{A}^*$ and $J \in \mathcal{A}$ it is always possible to find a sequence of sets I_s as above with $I_1 = I$ and $I_2 = J$. This implies the first assertion. In addition, $\pi_{J\setminus I} P$ is uniform and

$\pi_J P$ is the product of $\pi_{J\cap I} P$ with this uniform pm, using (15). Therefore, denoting by K the set of coloops of $\mathcal{A}^*$, the marginal $\pi_J P$ is the product of $\pi_{J\cap K} P$ with the uniform pm on $X_{J\setminus K}$. If $\mathcal{A}^*$ has no coloop $\pi_J P$ is uniform. ∎

Example 1: Let $N = \{1, 2, 3, 4\}$ and $\mathcal{A}$ consist of the sets $\{1, 2\}$, $\{1, 3\}$ and $\{4\}$. The family $\mathcal{A}^*$ has the coloop 1. All state spaces are assumed to equal $\{0, 1\}$. The pm P sitting uniformly on $(0, 0, 0, 0)$ and $(0, 1, 1, 1)$, and the pm R uniform on $\{0\} \times X_2 \times X_3 \times X_4$ have the same marginals to any $J \in \mathcal{A}$, and R is $\mathcal{A}$-factorizable. If follows that $D(P\|\mathcal{E}_{N,\mathcal{A}})$ equals $D(P\|R) = 2 \ln 2$. Since the marginal $\pi_{\{1,2\}} P$ is not uniform the second assertion of Theorem 2 fails without the exclusion of coloops.

Remark 4: When examining eq. (14) the two decomposition lemmas can be used to remove loops and coloops as follows. By Corollaries 1 and 3 with $K = \bigcup \mathcal{A}$, it is not difficult to see that a pm P satisfies (14) if and only if $H(\pi_K P)$ equals $H(P)$ and $D(\pi_K P\|\mathcal{E}_{K,\mathcal{A}})$ equals $\ln \min_{I\in\mathcal{A}} |X_{K\setminus I}|$, which is the upper bound of Corollary 3 with $(K, \mathcal{A})$ in the role of $(N, \mathcal{A})$. By Lemma 4, denoting J the set of coloops of $\mathcal{A}$ and $K = N \setminus J$, a pm P satisfies (14) if and only if for each x_J that belongs to the support of $\pi_J P$

$$D(P(\cdot|x_J)\|\mathcal{E}_{K,\mathsf{re}_K \mathcal{A}}) = \ln \min_{I\in \mathsf{re}_K \mathcal{A}} |X_{K\setminus I}|$$

which is the upper bound of Corollary 3 with $(K, \mathsf{re}_K \mathcal{A})$ replacing $(N, \mathcal{A})$.

IV. Matroid representations by partitions

In this section, the role of $\mathcal{A}$ is played by the family of bases $\mathcal{B}$ of a matroid with the ground set N. Let r denote the rank function of $\mathcal{B}$

$$r(K) = \max_{I\in\mathcal{B}} |I \cap K| , \qquad K \subseteq N . \qquad (16)$$

A base $I \in \mathcal{B}$ is characterized by $|I| = r(I) = r(N)$. A subset I of N is covered by a base if and only if $r(I) = |I|$ in which case it is called independent. The remaining subsets of N are dependent in the matroid. The inclusion minimal dependent sets are called circuits. Backgrounds on the matroid theory are collected in [18], [19].

A. Definition and basic properties

By [13, Definition 1], a matroid over N is *partition representable of the degree* $d \geqslant 2$ if there exist partitions ξ_i, $i \in N$, of a finite set Ω with $d^{r(N)}$ elements such that for all $I \subseteq N$ the meet of ξ_i, $i \in I$, has $d^{r(I)}$ blocks of the same cardinality. The collection $(\xi_i)_{i\in N}$ is called a partition representation of the matroid of the degree d.

Such a partition representation gives rise to the pm P on the product $X = \prod_{i\in N} X_i$ where X_i is the set of blocks of ξ_i

$$P(x) = \begin{cases} d^{-r(N)} , & \bigcap_{i\in N} \pi_i x \neq \emptyset , \\ 0 , & \text{otherwise} , \end{cases} \qquad x \in X . \qquad (17)$$

The above intersection is a singleton if nonempty. It is not difficult to see that

$$\pi_I P(x_I) = d^{-r(I)} , \qquad x_I \in \mathsf{s}(\pi_I P) , \, I \subseteq N . \qquad (18)$$

This means that $\mathsf{s}(\pi_I P)$ has the cardinality $d^{r(I)}$ and $\pi_I P$ is uniform on its support.

Remark 5: For a pm P on the product of finite sets X_i, $i \in N$, let ξ_i be the partition of $\Omega = \mathsf{s}(P)$ having x, y in the same block if and only if $\pi_i x = \pi_i y$. If P satisfies (18) with some $d \geqslant 2$ and the rank function r of a matroid then it is easy to see that $(\xi_i)_{i\in N}$ is a partition representation of the matroid of the degree d.

Remark 6: The partition representations are very closely related to the ideal secret sharing schemes, or to the almost affine codes, or to probabilistic representations of matroid-induced conditional independence relations. These relations are discussed in detail in [13, Section 5] and [16, Section 4].

The equations (18) imply that $H(\pi_I P) = r(I) \ln d$, $I \subseteq N$, thus the entropies of marginals are proportional to the ranks with the multiplicative factor $\ln d$. The following assertion shows that this implication can be reversed under some assumptions. Recall that a matroid is connected if $r(N)$ equals $r(I) + r(N \setminus I)$ only when I is $\emptyset$ or N. A nonconnected matroid decomposes into connected components.

Theorem 3: For a matroid over N with the rank function r, $d > 1$ and pm P on X, if $|\mathsf{s}(\pi_i P)| \leqslant d$ for each i from every connected component of the matroid that has the rank one and $H(\pi_I P) = r(I) \ln d$, $I \subseteq N$, then (18) holds. If also $r(N) > 0$ then d is integer.

This assertion is a variation on [12, Theorem]. For reader's convenience a proof is presented in Appendix.

B. Attainment of the upper bound via representations

In this and the following subsection it is assumed that the state spaces X_i, $i \in N$, have the same cardinality $d \geqslant 2$.

Remark 7: If a pm P corresponds to a partition representation of a matroid $(N, \mathcal{B})$ as in Remark 5, thus satisfies (18), then for every $I \in \mathcal{B}$ the marginal $\pi_I P$ is uniform on X_I and $H(P)$ is equal to $H(\pi_I P)$. By Lemma 5, $D(P\|\mathcal{E}_{N,\mathcal{B}})$ equals $[\,|N| - r(N)\,] \ln d$ which is the upper bound of Corollary 3. Such pm's however do not exhaust all maximizers of $D(\cdot\|\mathcal{E}_{N,\mathcal{B}})$ that attain the bound. In fact, consider another matroid $(N, \mathcal{B}')$ such that $\mathcal{B}$ is a subfamily of $\mathcal{B}'$. If a pm Q corresponds to a representation of $(N, \mathcal{B}')$ as in Remark 5 then, by the same argumentation as above, $D(Q\|\mathcal{E}_{N,\mathcal{B}})$ attains the upper bound $[\,|N|-r(N)\,] \ln d$. This situation is illustrated by the following example.

Example 2: Let $N = \{1, 2, 3, 4\}$,

$$\mathcal{B} = \{\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$$

and $\mathcal{B}' \supseteq \mathcal{B}$ contain in addition the set $\{3, 4\}$. Let partition representations have the degree $d = 2$ and all state spaces consist of 0 and 1. The pm P sitting uniformly on the four elements $(0, 0, 0, 0)$, $(1, 1, 0, 0)$, $(0, 0, 1, 1)$ and $(1, 1, 1, 1)$ of X corresponds to a representation of $(N, \mathcal{B})$ while the pm Q sitting uniformly on $(0, 0, 0, 0)$, $(0, 0, 1, 1)$, $(1, 1, 0, 1)$ and $(1, 1, 1, 0)$ to a representation of $(N, \mathcal{B}')$. In both cases, $D(P\|\mathcal{E}_{N,\mathcal{B}})$ and $D(Q\|\mathcal{E}_{N,\mathcal{B}})$ are equal to $2 \ln 2$, the upper bound on $D(\cdot\|\mathcal{E}_{N,\mathcal{B}})$ of Corollary 3.

Remark 8: If a pm P corresponds to a partition representation of a matroid $(N,\mathcal{B})$ as in Remark 5 then for every set $J \subseteq N$ the marginal pm $\pi_J P$ corresponds to a partition representation of the restriction of the matroid to J. By Remark 7, $D(\pi_J P \| \mathcal{E}_{J,\mathsf{re}_J\,\mathcal{B}})$ equals $[\,|J|-r(J)\,]\,\ln d$, which is the upper bound on $D(\cdot\|\mathcal{E}_{J,\mathsf{re}_J\,\mathcal{B}})$ resulting from Corollary 3.

C. Tightness of the upper bound for matroids

By Theorem 2, if a matroid $(N,\mathcal{B})$ has no coloop and the function $D(\cdot\|\mathcal{E}_{N,\mathcal{B}})$ attains the upper bound $[\,|N|-r(N)\,]\,\ln d$ at some pm P then $H(P)=H(\pi_I P)=r(N)\ln d$ and $\pi_I P$ is uniform for all $I\in\mathcal{B}$. In general, this conclusion is weaker than (18), as visible in Example 2 for the marginal $\pi_{\{3,4\}}Q$.

However, in the case of a uniform matroid without coloops, when $\mathcal{B}$ consists of all subsets of N of the cardinality t satisfying $0\leqslant t<|N|$, a pm P on X has $D(P\|\mathcal{E}_{N,\mathcal{B}})$ equal to $[\,|N|-t\,]\,\ln d$ if and only if (18) holds. Thus, by Remark 5, the upper bound is attained if and only if the uniform matroid is partition representable of the degree d. Such representations of uniform matroids correspond to the families of $|N|-t$ orthogonal Latin hypercubes of the size d [13, Example 1.5]. In particular, if $|N|=4$ and $t=2$ then the corresponding matroid, $U_{2,4}$, is partition representable of the degree d if and only if two orthogonal Latin squares of the size d exist. Hence, the bound is attainable if and only if $d\geqslant 2$ is different from 2 and 6. In the two excluded cases the maximum of $D(\cdot\|\mathcal{E}_{N,\mathcal{B}})$ is not known.

For a general matroid, even a uniform one, a necessary and sufficient condition for $D(\cdot\|\mathcal{E}_{N,\mathcal{B}})$ to attain the bound resulting from Corollary 3 remains elusive. The exclusion of loops and coloops is not restrictive in the problem, due to Remark 4.

Excluding coloops, the following result reveals that when each marginal $\pi_I P$ of a pm P makes $D(\pi_I P\|\mathcal{E}_{I,\mathsf{re}_I\,\mathcal{A}})$ to attain the corresponding upper bound of Corollary 3 then matroidal structures of $\mathcal{A}$ and P are unavoidable. In particular, the assertion of Remark 8 is reversed.

Theorem 4: Let N be a finite set, $\mathcal{A}$ a nonempty family of subsets of N and r the rank function of $\mathcal{A}$ given as in (16). Let $t=r(N)$, $\mathcal{A}^*=\{I\in\mathcal{A}\colon |I|=t\}$ have no coloop and all the state spaces X_i, $i\in N$, have the same cardinality $d\geqslant 2$. If a pm P on X satisfies

$$D(\pi_J P\|\mathcal{E}_{J,\mathsf{re}_J\,\mathcal{A}})=[\,|J|-r(J)\,]\,\ln d\,,\qquad J\subseteq N\,,\qquad (19)$$

then the inclusion maximal sets from $\mathcal{A}$ form the family of bases of a matroid and (18) holds for P.

Proof: Theorem 2 applies to conclude that the marginal $\pi_I P$ is uniform on X_I for $I\in\mathcal{A}$. Thus, the equations in (18) hold when I is covered by some set from $\mathcal{A}$.

Assume $I\subseteq N$ is covered by no set from $\mathcal{A}$. Any inclusion maximal set K from $\mathsf{re}_I\,\mathcal{A}$ is properly contained in I. For such K and $i\in I\setminus K$ no set from $\mathsf{re}_I\,\mathcal{A}$ covers $K\cup i$, and thus there exists an inclusion minimal subset L of $K\cup i$ that is covered by no set from $\mathsf{re}_I\,\mathcal{A}$. Then, i belongs to L and $\mathsf{re}_L\,\mathcal{A}$ contains all sets $L\setminus\ell$ with $\ell\in L$, but not L. By (19), Theorem 2 applies to $(L,\mathsf{re}_L\,\mathcal{A})$ and the marginal $\pi_L P$, and implies that $\pi_{L\setminus i}P$ has the same entropy as $\pi_L P$.

Therefore, $H(\pi_{K\cup i}P)$ equals $H(\pi_K P)$. Since $i\in I\setminus K$ was arbitrary $H(\pi_I P)$ equals $H(\pi_K P)$. It follows that $\mathsf{s}(\pi_I P)$ has $|\mathsf{s}(\pi_K P)|=d^{|K|}$ elements and $\pi_I P$ is uniform on its support, using that K is covered by some set from $\mathcal{A}$.

Hence, if $I\subseteq N$ is covered by no set from $\mathcal{A}$ then all the inclusion maximal sets from $\mathsf{re}_I\,\mathcal{A}$ have the same cardinality $\ln|\mathsf{s}(\pi_I P)|/\ln d$, equal to $r(I)$ by the definition of the rank function. By [18, Theorem 2, p. 14], the inclusion maximal sets of $\mathcal{A}$ form a family of the bases of a matroid. The rank function of the matroid is r. For every $I\subseteq N$ it follows that $\mathsf{s}(\pi_I P)$ has $d^{r(I)}$ elements of the same $\pi_I P$-probability which implies the validity of (18) for P. ■

Remark 9: As a consequence, the matroid representations of a matroid without coloops, disguised into the pm's P satisfying (18) as in Remark 5, can be equivalently defined by requiring each marginal $\pi_J P$ to have the divergence $D(\pi_J P\|\mathcal{E}_{J,\mathsf{re}_J\,\mathcal{A}})$ that attains the corresponding upper bound of Corollary 3.

V. Discussion

Assume $x^{(1)},\dots,x^{(T)}$ is a sequence of elements from X and Q a pm from the family $\mathcal{E}_{N,\mathcal{A}}$. The product $\Pi_{t=1}^{T}Q(x^{(t)})$ characterizes numerically a relation between the sequence and pm. Statisticians speak about a sequence of i.i.d. observations from an unknown pm belonging to the hierarchical log-linear model and a likelihood.

The maximum likelihood (ML) principle considers a maximizer of the product $\Pi_{t=1}^{T}Q(x^{(t)})$ subject to $Q\in\mathcal{E}_{N,\mathcal{A}}$ for an *ML estimate*. Actually, no maximizer may exist but it exists and is unique if maximizing alternatively over $\mathsf{cl}(\mathcal{E}_{N,\mathcal{A}})$. This follows from a general theory of ML estimation in exponential families, see [7] for recent revisions and completions.

The empirical distribution P of the sequence is obtained by demanding $T\cdot P(x)$ to be equal to the number of $t's$ between 1 and T such that $x^{(t)}=x$, for $x\in X$. Since

$$D(P\|Q)=-H(P)-\tfrac{1}{T}\,\ln\textstyle\prod_{x\in\mathsf{s}(P)}Q(x)^{T\cdot P(x)}$$

the maximization considered in ML principle is directly related to the minimization of $D(P\|\cdot)$ over $\mathcal{E}_{N,\mathcal{A}}$. Therefore, the number $D(P\|\mathcal{E}_{N,\mathcal{A}})$ evaluates how the ML principle works on the sequence and family. When this divergence is small, the model fits the data, in a statistical language. If the empirical distribution however maximizes $D(\cdot\|\mathcal{E}_{N,\mathcal{A}})$ then the observations are the most unfavorable w.r.t. the given model, from the viewpoint of the ML principle.

The results presented in the previous section establish a new link between the hierarchical log-linear models, ML estimation and ideal secret sharing, and demonstrate that a statistically unfavorable situation can have a distinguished cryptographic content. For example, assume a matroid $(N,\mathcal{B})$ is uniform of rank $t<|N|$ and P is its partition representation of the degree $d\geqslant 2$, in the sense of Remark 5 with all the state spaces of the cardinality d. If a sequence $x^{(1)},\dots,x^{(T)}$ lists all elements of $\mathsf{s}(P)$ without repetitions, $T=d^{t}$, then the empirical distribution of the sequence equals P and is the most unfavorable with respect to the model $\mathcal{E}_{N,\mathcal{B}}$. At the same time, P corresponds to a threshold scheme, a particularly simple instance of ideal secret sharing. For general matroids, such an interpretation can be based on Remark 9.

The cryptographic content becomes clear when a maximizer P satisfying (18) is interpreted as an ideal secret sharing scheme. This is outlined as follows. The elements of N are called the participants and an arbitrarily chosen i from N the dealer of a scheme. The elements of X_i are called the secrets. The aim is to choose a secret x_i and communicate to each participant $j \in N \setminus i$ a share $x_j \in X_j$ such that, pooling shares together, authorized groups of participants can recover the secret while the secret remains concealed from unauthorized groups of participants. Here, a group is authorized if it contains $I \setminus i$ for some circuit I containing i.

Sampling x from P, the coordinate $\pi_i x$ becomes the secret and $\pi_j x$ the share of the j-th participant. Each authorized group K can find the secret $\pi_i x$ uniquely from the pooled shares $\pi_K x$ because $H(\pi_{i\cup K} P)$ equals $H(\pi_K P)$. If K is not authorized then $H(\pi_{i\cup K} P) = H(\pi_i P) + H(\pi_K P)$, thus the scheme is perfect. Since $H(\pi_j P) = \ln d$ for all $j \in N$ it is even ideal.

Appendix

This section contains a proof of Theorem 3. It is shown first that for any circuit L with $|L| \geqslant 3$

$$\pi_i P(x_i) = d^{-1}\,, \qquad i \in L\,,\, x_i \in \mathsf{s}(\pi_i P)\,. \tag{20}$$

In fact, for $i, j \in L$ different $\pi_{i\cup j} P$ is the product if its marginals because $H(\pi_{ij} P) = H(\pi_i P) + H(\pi_j P)$ follows from the assumptions on entropies. Hence, for $x_i \in \mathsf{s}(\pi_i P)$ and $x_j \in \mathsf{s}(\pi_j P)$ there exists $x_L \in \mathsf{s}(\pi_L P)$ that projects on x_i and x_j. In turn, for $K = L \setminus (i \cup j)$ the assumptions on entropies imply that

$$\pi_i P(x_i) \cdot \pi_K P(\pi_K x_L) = \pi_{i\cup K} P(\pi_{i\cup K} x_L) = \pi_L P(x_L)\,.$$

By symmetry, the equalities hold also when i is replaced by j. Comparing the two pairs of equalities, $\pi_i P(x_i) = \pi_j P(x_j)$ for any $x_i \in \mathsf{s}(\pi_i P)$ and $x_j \in \mathsf{s}(\pi_j P)$. This observation combines with $H(\pi_i P) = \ln d$ to conclude that $d \geqslant 2$ is integer, the support of $\pi_i P$ has the cardinality d, and this marginal pm sits uniformly on it.

Let a connected component of the matroid has the ground set L. If $r(L) = 1$ then (20) is a consequence of $|\mathsf{s}(\pi_i P)| \leqslant d$ and $H(\pi_i P) = \ln d$, $i \in L$. If $r(L) \geqslant 2$ then L partitions into the blocks of mutually parallel elements. Any two elements of L that are not parallel are contained in a circuit of size at least three [19, Theorem 4.1.4]. Then, (20) follows from its starting version with circuits.

To summarize, (20) is true if L denotes the set of elements of N that are not the loops of the matroid. Hence, d is integer if not all elements are the loops, thus $r(N) > 1$.

It follows from the assumptions on entropies that if J is an independent set of the matroid then the marginal of P to J sits uniformly on $\mathsf{s}(\pi_J P) = \prod_{j\in J} \mathsf{s}(\pi_j P)$ which has $d^{|J|}$ elements. Since the rank of any subset I of N equals $r(J)$ for a maximal independent set J contained in I, the entropies of $\pi_I P$ and $\pi_J P$ coincide. Therefore, the two marginals sit uniformly on sets of the same cardinality, thus (18) holds.

Acknowledgment

I would like to express my gratitude to Thomas Kahle, Leipzig, for motivating discussions on the topic in the spring 2008. The final version is based on inspiring discussion with Nihat Ay and his students in October 2008, in Max Planck Institute for Mathematics in the Sciences, Leipzig.

References

[1] N. Ay: An information-geometric approach to a theory of pragmatic structuring. *Ann. Probab.* **30** (2002) 416–436.
[2] N. Ay and A. Knauf: Maximizing multi-information. *Kybernetika* **45** (2006) 517–538.
[3] O. Barndorff-Nielsen: *Information and Exponential Families in Statistical Theory.* Wiley, New York 1978.
[4] L.D. Brown: *Fundamentals of Statistical Exponential Families.* Inst. of Math. Statist. Lecture Notes – Monograph Series, Vol. 9, 1986.
[5] N. N. Chentsov: *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs, American Mathematical Society, Providence, R.I. 1982. (Russian original: Nauka, Moscow 1972.)
[6] I. Csiszár and F. Matúš: Information projections revisited. *IEEE Trans. Inform. Theory* **49** (2003) 1474–1490.
[7] I. Csiszár and F. Matúš: Generalized maximum likelihood estimates for exponential families. *Probability Theory and Related Fields* **141** (2008) 213–246.
[8] I. Csiszár and P.C. Shields: *Information Theory And Statistics: A Tutorial. Foundations and Trends in Communications and Information Theory*, Vol. 1, **4** (2004) 417-528.
[9] D. Geiger, Ch. Meek and B. Sturmfels: On the toric algebra of graphical models. *Ann. Statist.* **34** (2006) 1463–1492.
[10] S. J. Haberman: *The Analysis of Frequency Data.* University of Chicago Press, Chicago 1974.
[11] S. Della Pietra, V. Della Pietra, and J. Lafferty: Inducing features of random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **19** (1997) 380–393.
[12] F. Matúš: Probabilistic conditional independence structures and matroid theory: background. *International Journal of General Systems* **22** (1994) 185–196.
[13] F. Matúš: Matroid representations by partitions. *Discrete Mathematics* **203** (1999) 169–194.
[14] F. Matúš and N. Ay: On maximization of the information divergence from an exponential family. In: *Proc. WUPES'03* (J. Vejnarová, ed.), University of Economics, Prague 2003, pp. 199–204.
[15] F. Matúš: Maximization of information divergences from binary i.i.d. sequences. In: *Proc. IPMU 2004*, Perugia 2004, Vol. 2, pp. 1303–1306.
[16] F. Matúš: Two constructions on limits of entropy functions. *IEEE Trans. Inform. Theory* **53** (2007) 320–330.
[17] F. Matúš: Optimality conditions for maximizers of the information divergence from an exponential family. *Kybernetika* **43** (2007) 731–746.
[18] D.J.A. Welsh: *Matroid Theory.* Academic Press, London 1976.
[19] J.G. Oxley: *Matroid Theory.* Oxford University Press, Oxford, New York, Tokyo 1992.
[20] R.W. Yeung: *A First Course in Information Theory*. Kluwer Academic/Plenum Publishers, New York 2002.