BOUNDING $\bar{d}$-DISTANCE BY INFORMATIONAL DIVERGENCE: A METHOD TO PROVE MEASURE CONCENTRATION[1]

By K. Marton

Mathematical Institute of the Hungarian Academy of Sciences

There is a simple inequality by Pinsker between variational distance and informational divergence of probability measures defined on arbitrary probability spaces. We shall consider probability measures on sequences taken from countable alphabets, and derive, from Pinsker's inequality, bounds on the $\bar{d}$-distance by informational divergence. Such bounds can be used to prove the "concentration of measure" phenomenon for some nonproduct distributions.

1. Introduction. Statement of the results. Let $\mathscr{X}$ be a countable set and let q^n and p^n be two probability measures on $\mathscr{X}^n$.

We denote by x^n the sequence $(x_1, x_2, \ldots, x_n) \in \mathscr{X}^n$ and by $\bar{d}$ the normed Hamming distance on $\mathscr{X}^n \times \mathscr{X}^n$:

$$\bar{d}(x^n, y^n) = n^{-1} \sum_{i=1}^{n} d(x_i, y_i),$$

where $d(x_i, y_i) = 1$ if $x_i \neq y_i$ and 0 otherwise.

The $\bar{d}$-distance between p^n and q^n is

$$\bar{d}(p^n, q^n) = \min \mathrm{E}\, \bar{d}(\hat{X}^n, X^n),$$

where the min is taken over all joint distributions with marginals $p^n = \mathrm{dist}\, \hat{X}^n$ and $q^n = \mathrm{dist}\, X^n$. ($\hat{X}^n$ and X^n denote random sequences with values in $\mathscr{X}^n$.) The informational divergence of p^n with respect to q^n is

$$D(p^n \| q^n) = \sum_{x^n \in \mathscr{X}^n} p(x^n) \log \frac{p(x^n)}{q(x^n)}.$$

[The right-hand side may be ∞ even if $q(x^n) > 0$ for all x^n with $p(x^n) > 0$.]

There is a simple but powerful inequality by Pinsker between variational distance and informational divergence [Pinsker (1964); Csiszár and Körner (1981)]. In Marton (1986) this inequality was generalized to one between $\bar{d}(p^n, q^n)$ and $n^{-1}D(p^n \| q^n)$ in the case when q is i.i.d. (i.e., the product of n identical distributions).

The aim of this paper is to generalize this inequality for certain nonproduct distributions q^n. More precisely, for some nonproduct distributions q^n we shall prove an inequality of the form

$$(1.1) \qquad \bar{d}(p^n, q^n) \le c\left[\frac{1}{2n} D(p^n \| q^n)\right]^{1/2}$$

for all n and all distributions p^n and $\mathcal{X}^n$. Such bounds yield a proof of the "concentration of measure" phenomenon for some nonindependent processes $\{X_i\}$ or distributions q. If $A \subset \mathcal{X}^n$, then let $[A]_\varepsilon$ denote the ε-neighborhood of A:

$$[A]_\varepsilon = \left\{y^n \in \mathcal{X}^n : \bar{d}(x^n, y^n) \le \varepsilon \text{ for some } x^n \in A\right\}.$$

It is known that if q is i.i.d., then it has the measure-concentration property. This means the following (in the i.i.d case): if $A \subset \mathcal{X}^n$, then

$$(1.2) \qquad q^n([A]_\varepsilon) \ge 1 - \exp\left[-2n\left(\varepsilon - \sqrt{\frac{1}{2n}\log\frac{1}{q^n(A)}}\right)^2\right]$$

provided $\varepsilon \ge [(2n)^{-1}\log(1/q^n(A))]^{1/2}$ [cf. McDiarmid (1989) and Talagrand (1995)]. We shall prove a similar inequality for any process $\{X_i\}$ or distribution q satisfying inequality (1.1) (Proposition 4).

Measure-concentration inequalities can be used in various problems of probability theory. A great variety of applications is given in a recent paper by Talagrand (1995). Talagrand only considers i.i.d. processes, but some of the problems he treated are meaningful also for more general processes [e.g., the bin-packing problem and the problem of the longest increasing subsequence; see Talagrand (1995)]. Let us mention that in Talagrand (1995) a more powerful inequality than (1.1) is used, but only for the case of product measures. In a forthcoming paper [Marton (1995a)] we shall extend the method of this work to generalize Talagrand's inequality to the Markov case.

Now we proceed to the formulation of the results. If p and r are probability distributions on $\mathcal{X}$, then $|p - r|$ will denote their variational distance (divided by 2). $\bar{d}(p^n, q^n)$ is a natural generalization of $|p - r|$, since

$$|p - r| = \min \Pr\{\hat{X} \ne X\},$$

where the min is taken over all joint distributions $\mathrm{dist}(\hat{X}, X)$ having marginals $p = \mathrm{dist}\,\hat{X}$ and $r = \mathrm{dist}\,X$. Let us first recall Pinsker's inequality.

FACT.

$$|p - r| \le \left[\tfrac{1}{2}D(p\|r)\right]^{1/2}$$

[cf. Pinsker (1964), where this inequality was proved with a worse constant, and Csiszár and Körner (1981)].

Let q^n be a Markov measure on $\mathscr{X}^n$, that is, $q(x^n) = q_1(x_1)\prod_{i=2}^n q_i(x_i|x_{i-1})$. We denote by $q_i(\cdot)$ the ith marginal distribution of q^n:

$$q_i(x_i) = \sum_{x^{i-1}\in\mathscr{X}^{i-1}} q_1(x_1) \prod_{j=2}^i q_j(x_j|x_{j-1}).$$

PROPOSITION 1.   *Let q^n be a Markov measure on $\mathscr{X}^n$ and assume that*

(1.3) $$\max_i \sup_{\hat{x},\, x\in\mathscr{X}} |q_i(\cdot\,|\,\hat{x}) - q_i(\cdot\,|\,x)| = 1 - a, \qquad a > 0.$$

Then for any probability measure p^n on $\mathscr{X}^n$,

(1.4) $$\bar{d}(p^n, q^n) \le \frac{1}{a}\left[\frac{1}{2n}D(p^n\|q^n)\right]^{1/2}.$$

Inequality (1.4) implies for a product measure q^n,

$$\bar{d}(p^n, q^n) \le \left[\frac{1}{2n}D(p^n\|q^n)\right]^{1/2}.$$

In Marton (1986) this was proved, somewhat awkwardly, with a factor 2 on the right-hand side.

Let $\mathscr{X}^{\mathbb{Z}}$ denote the set of doubly infinite sequences taken from $\mathscr{X}$. If

$$(\cdots x_{-1}, x_0, x_1, x_2 \cdots) \in \mathscr{X}^{\mathbb{Z}},$$

then we denote by x_l^k $(l < k)$ the subsequence $(x_{l+1}, x_{l+2}, \ldots, x_k)$. The index l is omitted when 0, that is, $x^k = (x_1, x_2, \ldots, x_k)$. If q is a stationary probability measure on $\mathscr{X}^{\mathbb{Z}}$, that is, q is the distribution of some stationary process $\{X_i\}_{i=-\infty}^{\infty}$, taking values in $\mathscr{X}$, then we use the notation $q^n = \text{dist}\, X^n$ and

$$q(\cdot|x_{-N}^0) = \text{dist}(X_1|X_{-N}^0 = x_{-N}^0).$$

THEOREM 2.   *Let q be a stationary measure on $\mathscr{X}^{\mathbb{Z}}$, and put*

$$\gamma_k = \sup_N \sup_{x_{-N}^0,\, y_{-N}^0:\, y_{-k}^0 = x_{-k}^0} |q(\cdot\,|\,x_{-N}^0) - q(\cdot\,|\,y_{-N}^0)|.$$

If $\sum_{k=1}^\infty \gamma_k = 1 - a$, $a > 0$, then for any n and any probability measure p^n on $\mathscr{X}^n$,

(1.5) $$\bar{d}(p^n, q^n) \le \frac{1}{a}\left[\frac{1}{2n}D(p^n\|q^n)\right]^{1/2}.$$

Note that the condition $\sum_{k=1}^\infty \gamma_k < 1$ is a very strong mixing condition, much stronger than ϕ-mixing.

Condition (1.3) may not hold for a segment of a stationary mixing Markov chain. Still we have the following bound. The condition in the next proposition holds automatically if q is mixing and $\mathscr{X}$ is finite; for a countable alphabet it is called Doeblin's condition.

PROPOSITION 3.   *Let $q = \text{dist}\{X_i\}_{i=-\infty}^{\infty}$ be a stationary Markov measure on $\mathscr{X}^{\mathbb{Z}}$ and assume that for some k,*

(1.6) $$\sup_{x,\,y\in\mathscr{X}} \big|\text{dist}(X_k|X_0 = x) - \text{dist}(X_k|X_0 = y)\big| = 1 - a$$

with $a > 0$. Then for $n = tk$ and any distribution p^n on $\mathscr{X}^n$,

$$\bar{d}(p^n, q^n) \le \frac{k^{3/2}}{a}\left[\frac{2}{n}D(p^n\|q^n)\right]^{1/2}.$$

(Unfortunately, the bound we get from this proposition for $k = 1$ is twice the bound of Proposition 1.)

We state the measure-concentration inequality that follows from (1.1) in a symmetric form.

PROPOSITION 4.   *If q satisfies (1.1) for all n and all distributions p^n on $\mathscr{X}^n$, then for any sets, $A, B \subset \mathscr{X}^n$ the Hamming distance between A and B satisfies*

$$\bar{d}(A, B) \le c\left[\left(\frac{1}{2n}\log\frac{1}{q^n(A)}\right)^{1/2} + \left(\frac{1}{2n}\log\frac{1}{q^n(B)}\right)^{1/2}\right].$$

Proposition 4 implies (1.2) with ε replaced by ε/c. To see this, one has to take the complement of $[A]_\varepsilon$ as B.

Proposition 3 and Theorem 2 show that Markov processes with Doeblin's condition, as well as processes with a very fast and uniform decay of memory, satisfy the condition of the above proposition.

REMARK 1.   The concentration of measure phenomenon is a sharper form of the blowing-up property, first proved for i.i.d. processes in Ahlswede, Gács and Körner (1976). The process $\{X_i\}$ or the distribution q has the blowing-up property if for any $\varepsilon > 0$ there are $\delta > 0$ and n_0 such that for $n \ge n_0$ and all $A \subset \mathscr{X}^n$,

$$q^n(A) \ge \exp(-n\delta) \quad \Rightarrow \quad q^n([A]_\varepsilon) \ge 1 - \varepsilon.$$

In Marton and Shields (1994) the blowing-up property was explored in more detail and its connections with other ergodic properties were established. It seems that the class of processes having the blowing-up property is much wider than that of the processes satisfying an inequality of type (1.1). We shall come back to the characterization of processes having the blowing-up property in a forthcoming paper [Marton (1995b)]. The blowing-up property for mixing Markov chains with finite alphabet follows also from the results in Marton and Shields (1994) and Papamarcou and Shalaby (1993).

REMARK 2.   Talagrand (1995) proves (1.2) for a probability space $(\mathscr{X}, \mathscr{B}, q)$ obtained from a metric space $\mathscr{X}$ (instead of a countable alphabet). Our arguments, too, carry over to metric spaces with a measurable diagonal [cf.

Marton (1986)]. In this case, the set $[A]_\varepsilon$ may not be measurable; therefore, in (1.2) one has to consider the inner measure of $[A]_\varepsilon$ instead of $q^n([A]_\varepsilon)$. A way to circumvent this is to use the symmetric form given in Proposition 4.

2. Proofs of the results. Although Proposition 1 [or, more precisely, the corresponding inequality (1.2)] is known for the i.i.d. case, we separately describe the proof for the case when q^n is a product measure. The reason is that this is the simplest application of a coupling idea based on the successive use of Pinsker's inequality. Also, this is a very short proof of Proposition 1 for the case of product measures.

Moreover, we shall give a separate proof of Proposition 1, in spite of the fact that it is a special case of Theorem 2. Our aim is to make the ideas of this proof, and also of the proof of Theorem 2, clearer.

We shall use the notation

$$p(\hat{x}^{i-1}) = \Pr\{\hat{X}^{i-1} = \hat{x}^{i-1}\}, \qquad p_i(\cdot|\hat{x}^{i-1}) = \mathrm{dist}\left(\hat{X}_i \,|\, \hat{X}^{i-1} = \hat{x}^{i-1}\right)$$

and

$$\Delta_i(\hat{x}^{i-1}) = D\left(p_i(\cdot|\hat{x}^{i-1})\|q_i(\cdot|\hat{x}^{i-1})\right).$$

We have then the following identity which expresses $(1/n)D(p^n\|q^n)$ as an average of the quantities $\Delta_i(\hat{x}^{i-1})$:

(2.1) $$\frac{1}{n}\sum_{i=1}^{n}\sum_{\hat{x}^{i-1}\in\mathscr{X}^{i-1}} p(\hat{x}^{i-1})\Delta_i(\hat{x}^{i-1}) = \frac{1}{n}D(p^n\|q^n).$$

This identity is crucial in our proofs.

PROOF OF PROPOSITION 1 FOR THE INDEPENDENT CASE. We assume that q^n is the product of distributions q_i. Let $\hat{X}^n$ and X^n denote random sequences distributed according to p^n and q^n, respectively. Our goal is to define a joint distribution $\mathrm{dist}(\hat{X}^n, X^n)$ so that

$$\mathrm{E}\bar{d}(\hat{X}^n, X^n) = \frac{1}{n}\sum_{i=1}^{n}\Pr\{\hat{X}_i \neq X_i\}$$

is sufficiently small.

By independence, we have now

(2.2) $$\Delta_i(\hat{x}^{i-1}) = D\left(p_i(\cdot|\hat{x}^{i-1})\|q_i\right).$$

Now let us define a joint distribution $\mathrm{dist}(\hat{X}^n, X^n)$ by induction as follows. Assume that for some i, $1 \leq i \leq n$, $\mathrm{dist}(\hat{X}^{i-1}, X^{i-1})$ is already defined. Fix a pair of sequences $(\hat{x}^{i-1}, x^{i-1})$. We have to define the joint conditional distribution

$$\mathrm{dist}\left(\hat{X}_i, X_i|\hat{x}^{i-1}, x^{i-1}\right).$$

This distribution shall have marginals

$$\mathrm{dist}\left(\hat{X}_i | \hat{x}^{i-1}, x^{i-1}\right) = p_i(\cdot | \hat{x}^{i-1})$$

and

$$\mathrm{dist}\left(X_i | \hat{x}^{i-1}, x^{i-1}\right) = q_i.$$

By Pinsker's inequality, we can achieve

$$\Pr\left\{\hat{X}_i \neq X_i | \hat{X}^{i-1} = \hat{x}^{i-1}\right\} \leq \left[\tfrac{1}{2}\Delta_i(\hat{x}^{i-1})\right]^{1/2}$$

for all $\hat{x}^{i-1}$.

Now we use identity (2.1). Taking the average with respect to $\hat{x}^{i-1}$ and i, and using the concavity of the square-root function, we get

$$\bar{d}(p^n, q^n) \leq \left[\frac{1}{2n} \sum_{i=1}^n p(\hat{x}^{i-1})\Delta(\hat{x}^{i-1})\right]^{1/2} = \left[\frac{1}{2n}D(p^n\|q^n)\right]^{1/2}. \qquad \square$$

PROOF OF PROPOSITION 1. Let $\hat{X}^n$ and X^n denote random sequences distributed according to p^n and q^n, respectively. Our goal is to define a joint distribution $\mathrm{dist}(\hat{X}^n, X^n)$ so that

$$\mathrm{E}\bar{d}(\hat{X}^n, X^n) = \frac{1}{n} \sum_{i=1}^n \Pr\left\{\hat{X}_i \neq X_i\right\}$$

is sufficiently small.

We have now

(2.3) $$\Delta_i(\hat{x}^{i-1}) = D\left(p_i(\cdot | \hat{x}^{i-1})\|q_i(\cdot | \hat{x}_{i-1})\right).$$

Let us define a random sequence $(\hat{X}^n, Y^n)$ [or, more precisely, a joint distribution $\mathrm{dist}(\hat{X}^n, Y^n)$] as follows. The components Y_i, $i = 1, 2, \ldots, n$, shall be conditionally independent, given $\hat{X}^n$, and shall satisfy

(2.4) $$\mathrm{dist}\left(Y_i | \hat{X}^{i-1} = \hat{x}^{i-1}\right) = q_i(\cdot | \hat{x}_{i-1})$$

and

$$\Pr\left\{Y_i \neq \hat{X}_i | \hat{X}^{i-1} = \hat{x}^{i-1}\right\} = \left| p_i(\cdot | \hat{x}^{i-1}) - q_i(\cdot | \hat{x}_{i-1})\right|.$$

Applying Pinsker's inequality to the measures $p_i(\cdot | \hat{x}^{i-1})$ and $q_i(\cdot | \hat{x}_{i-1})$, we get

$$\Pr\left\{Y_i \neq \hat{X}_i | \hat{X}^{i-1} = \hat{x}^{i-1}\right\} \leq \left[\tfrac{1}{2}\Delta_i(\hat{x}^{i-1})\right]^{1/2}.$$

The above conditions do not uniquely determine the joint distribution of $(\hat{X}^n, Y^n)$, but it is obvious that the definition can be completed and we can have

$$\Pr\left\{Y_i \neq \hat{X}_i | \hat{X}^{i-1} = \hat{x}^{i-1}, Y^{i-1} = y^{i-1}\right\} \leq \left[\tfrac{1}{2}\Delta_i(\hat{x}^{i-1})\right]^{1/2}$$

for all $\hat{x}^{i-1}, y^{i-1}$.

Now let us define a joint distribution $\mathrm{dist}(\hat{X}^n, Y^n, X^n)$ by induction as follows. Assume that for some i, $1 \le i \le n$, $\mathrm{dist}(\hat{X}^{i-1}, Y^{i-1}, X^{i-1})$ is already defined. Fix a triple of sequences $(\hat{x}^{i-1}, y^{i-1}, x^{i-1})$. We have to define the joint conditional distribution

$$\mathrm{dist}\left(\hat{X}_i, Y_i, X_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1}\right).$$

This distribution shall have marginals

$$\mathrm{dist}\left(\hat{X}_i, Y_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1}\right) = \mathrm{dist}\left(\hat{X}_i, Y_i | \hat{x}^{i-1}\right)$$

and

$$\mathrm{dist}\left(X_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1}\right) = q_i(\cdot | x_{i-1}).$$

(Both distributions are given.) By Pinsker's inequality we can achieve

$$\Pr\left\{Y_i \ne X_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1}\right\} = \left|q_i(\cdot | \hat{x}_{i-1}) - q_i(\cdot | x_{i-1})\right|$$

and, consequently,

$$\Pr\left\{\hat{X}_i \ne X_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1}\right\} \le \left[\tfrac{1}{2}\Delta_i(\hat{x}^{i-1})\right]^{1/2} + \left|q_i(\cdot | \hat{x}_{i-1}) - q_i(\cdot | x_{i-1})\right|.$$

Taking the average with respect to $\hat{x}^{i-1}$ and i, and using (2.1) and the concavity of the square-root function, we get

$$\frac{1}{n} \sum_{i=1}^{n} \Pr\left\{\hat{X}_i \ne X_i\right\} \le \frac{1}{n} D(p^n \| q^n) + (1 - a)\frac{1}{n} \sum_{i=1}^{n} \Pr\left\{\hat{X}_{i-1} \ne X_{i-1}\right\},$$

where, by definition, $\Pr\{\hat{X}_0 \ne X_0\} = 0$. Equivalently,

$$a\frac{1}{n} \sum_{i=1}^{n} \Pr\left\{\hat{X}_i \ne X_i\right\} \le \left[\frac{1}{2n} D(p^n \| q^n)\right]^{1/2}. \qquad \square$$

PROOF OF THEOREM 2. Let $\hat{X}^n$ and X^n denote random sequences distributed according to p^n and q^n, respectively. We shall imitate the proof of Proposition 1.

Define the joint distribution $\mathrm{dist}(\hat{X}^n, Y^n)$ as in the proof of Proposition 1, replacing (2.4) by

$$\mathrm{dist}\left(Y_i | \hat{X}^{i-1} = \hat{x}^{i-1}\right) = q_i(\cdot | \hat{x}^{i-1}).$$

We have again

$$(2.5) \qquad \mathrm{E}\bar{d}(\hat{X}^n, Y^n) = \frac{1}{n} \sum_{i=1}^{n} \Pr\left\{\hat{X}_i \ne Y_i\right\} \le \left[\frac{1}{2n} D(p^n \| q^n)\right]^{1/2}.$$

Now define the joint distribution $\mathrm{dist}(\hat{X}^n, Y^n, X^n)$ by induction, similarly to the proof of Proposition 1. Given the triple of sequences $(\hat{x}^{i-1}, y^{i-1}, x^{i-1})$, we define $\mathrm{dist}(Y_i, X_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1})$ with marginals $q_i(\cdot | \hat{x}^{i-1})$ and $q_i(\cdot | x^{i-1})$ so as to minimize

$$\Pr\left\{Y_i \ne X_i | \hat{x}^{i-1}, y^{i-1}, x^{i-1}\right\} = \Pr\left\{Y_i \ne X_i | \hat{x}^{i-1}, x^{i-1}\right\}.$$

With the notation $\hat{x}_j^{i-1} = (\hat{x}_{j+1}, \hat{x}_{j+2}, \ldots, \hat{x}_{i-1})$, we shall have

$$\Pr\{Y_i \neq X_i\} = \sum_{j=1}^{i-1} \Pr\{\hat{X}_j \neq X_j, \hat{X}_j^{i-1} = X_j^{i-1}\}$$

$$\times \Pr\{Y_i \neq X_i | \hat{X}_j \neq X_j, \hat{X}_j^{i-1} = X_j^{i-1}\}$$

$$\leq \sum_{j=1}^{i-1} \Pr\{\hat{X}_j \neq X_j\}\gamma_{i-j}.$$

It follows that

$$\mathrm{E}\bar{d}(Y^n, X^n) \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \Pr\{\hat{X}_j \neq X_j\}\gamma_{i-j}$$

$$= \frac{1}{n} \sum_{j=1}^{n-1} \Pr\{\hat{X}_j \neq X_j\} \sum_{i=j+1}^{n} \gamma_{i-j} \leq (1-a)\mathrm{E}\bar{d}(\hat{X}^n, X^n).$$

This, together with (2.5), implies

$$a\,\mathrm{E}\bar{d}(\hat{X}^n, X^n) \leq \left[(2n)^{-1}D(p^n\|q^n)\right]^{1/2}. \qquad \square$$

PROOF OF PROPOSITION 3. Let $\hat{X}^n = \hat{X}^{tk}$ and $X^n = X^{tk}$ denote random sequences distributed according to p^n and q^n, respectively. We imitate the coupling construction of the proof of Proposition 1, but now we concatenate k-length blocks defined as

$$\left(\hat{X}_{(i-1)k}^{ik}, Y_{(i-1)k}^{ik}, X_{(i-1)k}^{ik}\right) = \left\{\left(\hat{X}_j, Y_j, X_j\right): (i-1)k < j \leq ik\right\}.$$

For a fixed i, $1 \leq i \leq t$, and $\hat{x}^{(i-1)k} \in \mathscr{X}^{(i-1)k}$, write

$$\Delta_i(\hat{x}^{(i-1)k}) = D\left(p_{(i-1)k}^{ik}(\cdot|\hat{x}^{(i-1)k})\|q_{(i-1)k}^{ik}(\cdot|\hat{x}^{(i-1)k})\right),$$

where, for example, $p_{(i-1)k}^{ik}(\cdot|\hat{x}^{(i-1)k}) = \mathrm{dist}(\hat{X}_{(i-1)k}^{ik}|\hat{X}^{(i-1)k} = \hat{x}^{(i-1)k})$. We have, instead of (2.1),

$$\frac{1}{t} \sum_{i=1}^{t} \sum_{\hat{x}^{(i-1)k}} p(\hat{x}^{(i-1)k})\Delta_i(\hat{x}^{(i-1)k}) = \frac{k}{n}D(p^n\|q^n).$$

Let us define the random sequence Y^{tk} so that the strings $Y_{(i-1)k}^{ik}$, $i = 1, \ldots, t$, shall be conditionally independent, given $\hat{X}^n$, and satisfy

$$\mathrm{dist}\left(Y_{(i-1)k}^{ik}|\hat{X}^{(i-1)k} = \hat{x}^{(i-1)k}\right) = q^k\left(\cdot|\hat{x}_{(i-1)k}\right).$$

The right-hand side here denotes $\mathrm{dist}(X_{(i-1)k}^{ik}|X_{(i-1)k} = \hat{x}_{(i-1)k})$, that is, $Y_{(i-1)k}^{ik}$ develops from the state $\hat{x}_{(i-1)k}$ the same way as $X_{(i-1)k}^{ik}$ would do. Moreover, we define the joint conditional distribution of $(\hat{X}_{(i-1)k}^{ik}, Y_{(i-1)k}^{ik})$, given $\hat{X}^{(i-1)k}$ and $Y^{(i-1)k}$ so as to achieve

$$\Pr\left\{Y_{(i-1)k}^{ik} \neq \hat{X}_{(i-1)k}^{ik}|\hat{X}^{(i-1)k} = \hat{x}^{(i-1)k}, Y^{(i-1)k} = y^{(i-1)k}\right\}$$

$$\leq \left[\tfrac{1}{2}\Delta(\hat{x}^{(i-1)k})\right]^{1/2}, \quad \text{for all } \hat{x}^{(i-1)k}, y^{(i-1)k}.$$

Next we extend the definition of $\text{dist}(\hat{X}^n, Y^n)$ to a definition of $\text{dist}(\hat{Y}^n, Y^n, X^n)$, using induction in i and bounding, at each step, the conditional expected d-distance

$$\text{E}\left\{\bar{d}\left(Y_{(i-1)k}^{ik}, X_{(i-1)k}^{ik}\right) \mid \hat{X}^{(i-1)k} = \hat{x}^{(i-1)k}, Y^{(i-1)k} = y^{(i-1)k}, X^{(i-1)k} = x^{(i-1)k}\right\},$$

given the marginals $q^k(\cdot | \hat{x}_{(i-1)k})$ and $q^k(\cdot | x_{(i-1)k})$. This conditional expectation can be made less than or equal to $(1 - a/k)$ in case $\hat{x}_{(i-1)k} \neq x_{(i-1)k}$, and 0 if $\hat{x}_{(i-1)k} = x_{(i-1)k}$.

Now consider, for a given i, the random triple $(\hat{X}^{(i-1)k}, Y^{(i-1)k}, X^{(i-1)k})$, the distribution of which has been constructed, step by step, subject to the above conditions. Observe that if a pair

$$\left(y_{(i-2)k}^{(i-1)k}, x_{(i-2)k}^{(i-1)k}\right)$$

has positive probability and

$$\bar{d}\left(y_{(i-2)k}^{(i-1)k}, x_{(i-2)k}^{(i-1)k}\right) < 1,$$

then $y_{(i-1)k} = x_{(i-1)k}$, since both Y and X develop according to the same Markovian q. It follows that, if for some realization

$$\left(\hat{x}^{(i-1)k}, y^{(i-1)k}, x^{(i-1)k}\right)$$

we have $\hat{x}_{(i-2)k}^{(i-1)k} = y_{(i-2)k}^{(i-1)k}$, and $\bar{d}(\hat{x}_{(i-2)k}^{(i-1)k}, x_{(i-2)k}^{(i-1)k}) < 1$, then $\hat{x}_{(i-1)k} = x_{(i-1)k}$. Therefore,

$$\bar{d}\left(Y_{(i-1)k}^{ik}, X_{(i-1)k}^{ik}\right) \leq \left(1 - \frac{a}{k}\right)\bar{d}\left(\hat{X}_{(i-2)k}^{(i-1)k}, X_{(i-2)k}^{(i-1)k}\right),$$

provided $\hat{X}_{(i-2)k}^{(i-1)k} = Y_{(i-2)k}^{(i-1)k}$. (We assume $\hat{X}_{-k}^0 = Y_{-k}^0 = X_{-k}^0$.) It follows that

$$\text{E}\bar{d}\left(Y_{(i-1)k}^{ik}, X_{(i-1)k}^{ik}\right) \leq \Pr\left\{\hat{X}_{(i-2)k}^{(i-1)k} \neq Y_{(i-2)k}^{(i-1)k}\right\}$$

$$+ \left(1 - \frac{a}{k}\right)\text{E}\bar{d}\left(\hat{X}_{(i-2)k}^{(i-1)k}, X_{(i-2)k}^{(i-1)k}\right)$$

and so

$$\text{E}\bar{d}\left(\hat{X}_{(i-1)k}^{ik}, X_{(i-1)k}^{ik}\right)$$

$$\leq \Pr\left\{\hat{X}_{(i-2)k}^{(i-1)k} \neq Y_{(i-2)k}^{(i-1)k}\right\}$$

$$+ \Pr\left\{\hat{X}_{(i-1)k}^{ik} \neq Y_{(i-1)k}^{ik}\right\} + \left(1 - \frac{a}{k}\right)\text{E}\bar{d}\left(\hat{X}_{(i-2)k}^{(i-1)k}, X_{(i-2)k}^{(i-1)k}\right).$$

Averaging for i we get

$$\frac{a}{k}\text{E}\bar{d}(X^{tk}, \hat{X}^{tk}) \leq \left[\frac{2k}{n}D(p^n \| q^n)\right]^{1/2}. \qquad \square$$

PROOF OF PROPOSITION 4.   Let $p^n = \text{dist } \hat{X}^n$ denote the distribution

$$p^n(x^n) = \begin{cases} q(x^n)/q^n(A), & x^n \in A, \\ 0, & \text{otherwise,} \end{cases}$$

and define r^n similarly, replacing A with B. We have then

$$\bar{d}(p^n, q^n) \le c \left[\frac{1}{2n} \log \frac{1}{q^n(A)} \right]^{1/2}$$

and

$$\bar{d}(r^n, q^n) \le c \left[\frac{1}{2n} \log \frac{1}{q^n(B)} \right]^{1/2}.$$

Therefore,

$$\bar{d}(A, B) \le \bar{d}(p^n, r^n) \le \bar{d}(p^n, q^n) + \bar{d}(r^n, q^n)$$

$$\le c \left[\left(\frac{1}{2n} \log \frac{1}{q^n(A)} \right)^{1/2} + \left(\frac{1}{2n} \log \frac{1}{q^n(B)} \right)^{1/2} \right]. \qquad \square$$

REFERENCES

AHLSWEDE, R., GÁCS, P. and KÖRNER, J. (1976). Bounds on conditional probabilities with applications in multi-user communication. *Z. Wahrsch. Verw. Gebiete* **34** 157–177.

CSISZÁR, I. and KÖRNER, J. (1981). *Information Theory*: *Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York.

MARTON, K. (1986). A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory* **IT-32** 445–446.

MARTON, K. (1995a). A concentration-of-measure inequality for contracting Markov chains. *Geometric and Functional Analysis*. To appear.

MARTON, K. (1995b). Processes having the blowing-up property. Unpublished manuscript.

MARTON, K. and SHIELDS, P. C. (1994). The positive divergence and blowing-up properties. *Israeli J. Math.* **86** 331–348.

MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics*. *London Mathematical Society Lecture Notes* (J. Simons, ed.) **141** 148–188. Cambridge Univ. Press, London.

PAPAMARCOU, A. and SHALABY, H. (1993). Error exponent for distributed detection of Markov sources. *IEEE Trans. Inform. Theory* **40** 397–408.

PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco.

TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publ. IHES*. **81** 73–205.

MATHEMATICAL INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
H-1364 POB 127
BUDAPEST
HUNGARY
E-MAIL: marton@math-inst.hu