ISRAEL JOURNAL OF MATHEMATICS, Vol. 58, No. 3, 1987

ON WEAK ASYMPTOTIC ISOMORPHY OF MEMORYLESS CORRELATED SOURCES

BY
K. MARTON
Mathematical Institute of the Hungarian Academy of Sciences, P.O.B. 127, H-1364, Hungary

ABSTRACT

Let $\{(X_i, Z_i)\}$ be an i.i.d. sequence of random pairs in a finite set $\mathscr{X} \times \mathscr{Z}$; we will call it a discrete memoryless stationary correlated (DMSC) source with generic distribution dist(X_1, Z_1). Two DMSC sources $\{(X_i, Z_i)\}$ and $\{(X_i', Z_i')\}$ are called asymptotically isomorphic in the weak sense if for every $\varepsilon > 0$ and sufficiently large n, there exists a joint distribution dist(X^n, Z^n, X'^n, Z'^n) of n-length blocks of the two sources such that

$$\frac{1}{n} H(X^n \mid X'^n) < \varepsilon, \qquad \frac{1}{n} H(Z^n \mid Z'^n) < \varepsilon,$$

$$\frac{1}{n} H(X'^n \mid X^n) < \varepsilon, \qquad \frac{1}{n} H(Z'^n \mid Z^n) < \varepsilon.$$

For single sources of equal entropy, McMillan's theorem implies asymptotic isomorphy in the sense suggested by this definition. For correlated sources, however, no nontrivial cases of weak asymptotic isomorphy are known. We show that some spectral properties of the generic distributions are invariant for weak asymptotic isomorphy, and these properties wholly determine the generic distribution in many cases.

§1. Introduction

Let (X, Z) be a random pair with values in the set $\mathscr{X} \times \mathscr{Z}$ ($\mathscr{X}$, $\mathscr{Z}$ finite sets). Denote by P and R the distributions of X resp. Z, and by V and V^* the conditional distributions of Z given X resp. X given Z:

$$P = \text{dist } X, \qquad R = \text{dist } Z,$$

$$V = (V(z \mid x) : (x, z) \in \mathscr{X} \times \mathscr{Z}) = \text{dist}\,(Z \mid X),$$

$$V^* = (V^*(x \mid z) : (x, z) \in \mathscr{X} \times \mathscr{Z}) = \text{dist}\,(X \mid Z).$$

For the matrix V, the rows are indexed by the elements of $\mathscr{X}$, and for V^*, by the

Received December 30, 1985 and in revised form May 6, 1986

elements of $\mathscr{X}$. Let $\mathscr{L}(X)$ and $\mathscr{L}(Z)$ be the real L_2-spaces over the (finite) probability spaces $(\mathscr{X}, P)$ resp. $(\mathscr{Z}, R)$. The transition matrices V and V^* define Markov operators $\mathscr{L}(Z) \to \mathscr{L}(X)$ resp. $\mathscr{L}(X) \to \mathscr{L}(Z)$, by setting

$$Vg(x) = \sum_z V(z \mid x) g(z), \qquad g \in \mathscr{L}(Z),$$

$$V^* f(z) = \sum_x V^*(x \mid z) f(x), \qquad f \in \mathscr{L}(X).$$

It is easy to see that the operators V and V^* are adjoint to each other. Thus the Markov operators defined by the matrix products

$$W = VV^* : \mathscr{L}(X) \to \mathscr{L}(X),$$

$$\tilde{W} = V^* V : \mathscr{L}(Z) \to \mathscr{L}(Z)$$

are self-adjoint and non-negative definite. I.e.

$$(Wf, g)_P = (f, Wg)_P \quad \text{and} \quad (Wf, f)_P \geqq 0, \qquad f, g \in \mathscr{L}(X),$$

and

$$(\tilde{W}f, g)_R = (f, \tilde{W}g)_R \quad \text{and} \quad (\tilde{W}f, f)_R \geqq 0, \qquad f, g \in \mathscr{L}(Z),$$

where, e.g., the index P indicates that the scalar product is taken with respect to P. It is clear that the functions $\equiv 1$ on $\mathscr{X}$ resp. $\mathscr{Z}$ are eigenfunctions of W resp. $\tilde{W}$, with eigenvalue 1 which is the largest eigenvalue of W and $\tilde{W}$, so the spectra of W and $\tilde{W}$ lie in the interval $[0, 1]$.

Define two stationary Markov chains $\{X(m)\}_{m=0}^{\infty}$ and $\{Z(m)\}_{m=0}^{\infty}$ with dist $X(0) = P$, dist $Z(0) = R$, and transition matrices W and $\tilde{W}$, respectively. Denote $H(P)$ or $H(X)$ the Shannon entropy of the random variable X, and for a random pair (X, Y), denote $I(X \wedge Y)$ the mutual information $I(X \wedge Y) = H(X) + H(Y) - H(X, Y)$.

With this notation, put

$$I_m = I(X(0) \wedge X(m)), \qquad \tilde{I}_m = I(Z(0) \wedge Z(m)),$$

and

$$\hat{I}_m = I(U \wedge X(m)) \qquad (m = 1, 2, \ldots),$$

where $(U, X(0), X(1), \ldots)$ is a Markov chain and $\text{dist}(U, X(0)) = \text{dist}(Z, X)$.

Our goal is to show that the numbers $H(P)$, $H(R)$, $\{I_m\}$, $\{\tilde{I}_m\}$, $\{\hat{I}_m\}$ reveal much about the structure of the joint distribution $\text{dist}(X, Z)$, and in fact, wholly determine it in many cases. The motivation for studying these numbers

is the fact that they are invariant for weak asymptotic isomorphy of discrete memoryless stationary correlated (= DMSC) sources (cf.) Lemma 4.1 in [7]).

DEFINITION. The DMSC sources $\{(X_i, Z_i)\}$, $\{(X_i', Z_i')\}$ are asymptotically isomorphic in the weak sense if for any $\varepsilon > 0$ and sufficiently large n, there exists a joint distribution† $\mathrm{dist}(X^n, X'^n, Z^n, Z'^n)$ satisfying

$$\frac{1}{n} H(X^n \mid X'^n) < \varepsilon, \qquad \frac{1}{n} H(Z^n \mid Z'^n) < \varepsilon,$$

$$\frac{1}{n} H(X'^n \mid X^n) < \varepsilon, \qquad \frac{1}{n} H(Z'^n \mid Z^n) < \varepsilon.$$

For single sources $\{X_i\}$, $\{X_i'\}$ of the same entropy, McMillan's theorem implies that they are asymptotically isomorphic in the sense suggested by the above definition.

For correlated sources, the situation is quite different. No non-trivial cases of weak asymptotic isomorphy are known Moreover, adopting a somewhat more restrictive definition of asymptotic isomorphy, it turns out that DMSC sources cannot be asymptotically isomorphic in a non-trivial way. (See [8].) This suggests the conjecture that DMSC sources cannot be asymptotically isomorphic even in the weak sense, unless their generic distributions (i.e. $\mathrm{dist}(X_1, Z_1)$ and $\mathrm{dist}(X_1', Z_1')$ for the sources $\{(X_i, Z_i)\}$, $\{X_i', Z_i')\}$) are so.

Here we will show that some spectral properties of the matrix W are determined by $H(P)$ and the sequence $\{I_m\}$, and hence invariant for weak asymptotic isomorphy. Moreover, for a class of joint distributions, we will prove that if a joint distribution in this class has the same numbers $H(P)$, $H(R)$, $\{I_m\}$, $\{\tilde{I}_m\}$, $\{\hat{I}_m\}$ as another joint distribution (not necessarily in this class) then the two joint distributions are isomorphic in the following sense:

DEFINITION. Let Q and Q' be joint distributions on the sets $\mathcal{X} \times \mathcal{Z}$ resp. $\mathcal{X}' \times \mathcal{Z}'$ ($\mathcal{X}$, $\mathcal{Z}$, $\mathcal{X}'$, $\mathcal{Z}'$ are finite sets). Q and Q' are isomorphic if there exist bijections $\sigma : \mathcal{X} \to \mathcal{X}'$, $\tau : \mathcal{Z} \to \mathcal{Z}'$ such that

$$Q(x, z) = Q'(\sigma(x), \tau(x)) \qquad \text{for any } (x, z) \in \mathcal{X} \times \mathcal{Z}.$$

The study of the sequence $\{\hat{I}_m\}$ for dealing with a related problem was initiated by Thouvenot [10]. He considered DMSC sources that are full-entropy factors of some given DMSC source. Our Theorem 2 is a substantial

† X^n denotes $(X_1, \ldots, X_n)$.

generalization of Thouvenot's Proposition 1 in [10]. (The proof of his Proposition, however, contains a gap.) Let us note that we were not able to fully generalize Thouvenot's result (implied by Lemma 2 [10]) on the eigenvalues, to the case of weak asymptotic isomorphy.

Throughout the paper, we will consider pairs (X, Z) satisfying the following two conditions:

(i) 1 is a simple eigenvalue of W;

(ii) (X, Z) is non-singularly dependent (i.e. neither V nor V^* contain identical rows).

These assumptions are justified by

LEMMA 0 (proved in the Appendix). *Properties* (i) *and* (ii) *are invariant for weak asymptotic isomorphy.*

It is shown in [6] that property (ii) is equivalent to the condition that the eigenfunctions of W resp. $\tilde{W}$ pertaining to positive eigenvalues separate $\mathscr{X}$ resp. $\mathscr{Z}$.

§2. Statement of the results

Denote $\lambda_1, \ldots, \lambda_r$ the different eigenvalues of W lying in the interval $(0, 1)$, and let Λ be the set of those $\lambda_j \in \Lambda$ that admit no representation of the form

$$\lambda_j = \prod_{i=1}^{r} \lambda_i^{\alpha_{ji}}, \qquad \alpha_{ji} \geqq 0 \text{ rationals}, \qquad \sum_i \alpha_{ji} \geqq \tfrac{3}{2}.$$

Note that $\max_i \lambda_i \in \Lambda$, so $\Lambda \neq \varnothing$.

THEOREM 1. *The set* Λ *and the multiplicities of* $\lambda_i \in \Lambda$ *are determined by the sequence* $\{I_m\}$.

DEFINITION. The positive numbers $\mu_1, \ldots, \mu_s$ are called log-independent if for integers $n_1, \ldots, n_s$

$$\prod_{i=1}^{s} \mu_i^{n_i} = i \quad \text{imples } n_i = 0 \quad \text{for all } i.$$

NOTATION. Speaking of functions on $\mathscr{X}$ and $\mathscr{Z}$, orthogonality will be understood with respect to P resp. R. Denote $\{u_{ij}\}_{j=1,\ldots,s_i}$ a complete orthonormal set of λ_i-eigenfunctions of W $(i = 1, \ldots, r)$ and put

$$\tilde{u}_{ij} = \frac{1}{\sqrt{\lambda_i}} V^* u_{ij};$$

it is easy to see (or cf. [9, Translator's remarks to Chap. 9] that $\{\tilde{u}_{ij}\}_{j=1,\dots,s_i}$ is a complete orthonormal set of λ_i-eigenfunctions of $\tilde{W}$. Denote for $(x, y)\in\mathscr{X}^2$

$$F_{xy}(\zeta) = 1 + \sum_{i=1}^{r} \zeta_i \sum_{j=1}^{s_i} u_{ij}(x)u_{ij}(y), \qquad \zeta = (\zeta_1, \dots, \zeta_r)\in\mathbf{C}^r;$$

for $(x, z)\in\mathscr{X}\times\mathscr{Z}$,

$$\hat{F}_{xz}(\zeta) = 1 + \sum_{i=1}^{r} \zeta_i \sum_{j} u_{ij}(x)\tilde{u}_{ij}(z);$$

and for $(z, t)\in\mathscr{Z}^2$,

$$\tilde{F}_{zt}(\zeta) = 1 + \sum_{i=1}^{r} \zeta_i \sum_{j} \tilde{u}_{ij}(z)\tilde{u}_{ij}(t).$$

Denote

$$V^{*(m)} = V^*W^m = \operatorname{dist}(X(m) \mid U).$$

The meaning of the functions F_{xy}, $\hat{F}_{xz}$, $\tilde{F}_{zu}$ is clarified by

LEMMA 1. (i) $V^{*(m)}(x \mid z) = P(x)\hat{F}_{xz}(\lambda^{m+1/2})$, $(x, z)\in\mathscr{X}\times\mathscr{Z}$, *where* $\lambda^q = (\lambda_1^q, \dots, \lambda_r^q)$;

(ii) $W^m(y \mid x) = P(y)F_{xy}(\lambda^m)$, $(x, y)\in\mathscr{X}^2$;

(iii)
$$\hat{I}_m = I(U\wedge X(m))$$
$$= \sum_{x,z} P(x)R(z)\hat{F}_{xz}(\lambda^{m+1/2})\log \hat{F}_{xz}(\lambda^{m+1/2})$$
$$= \sum_{s=2}^{\infty} \frac{(-1)^s}{s(s-1)} \sum_{i_1+\cdots+i_r=s} \lambda_1^{i_1(m+1/2)} \cdots \lambda_r^{i_r(m+1/2)} \cdot \tag{2.1}$$
$$\cdot \sum_{xz} P(x)R(z) \left(\sum_j u_{1j}(x)\tilde{u}_{1j}(z)\right)^{i_1} \cdots \left(\sum_j u_{rj}(x)\tilde{u}_{rj}(z)\right)^{i_r}.$$

(iv)
$$I_m = I(X(0)\wedge X(m))$$
$$= \sum_{x,y} P(x)P(y)F_{xy}(\lambda^m)\log F_{xy}(\lambda^m)$$
$$= \sum_{s=2}^{\infty} \frac{(-1)^s}{s(s-1)} \sum_{i_1+\cdots+i_r=s} \lambda_1^{i_1 m} \cdots \lambda_r^{i_r m} \cdot \tag{2.2}$$
$$\cdot \sum_{xy} P(x)P(y) \left(\sum_j u_{1j}(x)u_{1j}(y)\right)^{i_1} \cdots \left(\sum_j u_{rj}(x)u_{rj}(y)\right)^{i_r}.$$

If $F(\zeta)$, $\zeta \in C^r$, is a linear function in r variables then define the sets $A(F)$, $B(F)$, $C(F)$ by

$$A(F) = \{(x, y) \in \mathcal{X}^2 : F_{xy} \equiv F\},$$

$$B(F) = \{(z, t) \in \mathcal{Z}^2 : \hat{F}_{zt} \equiv F\},$$

$$C(F) = \{(x, z) \in \mathcal{X} \times \mathcal{Z} : \hat{F}_{xz} \equiv F\}.$$

THEOREM 2. *Let W be non-singular, and the eigenvalues $\lambda_1, \ldots, \lambda_r$ log-independent. If for some* dist(X', Z')

$$I_m = I'_m, \quad \tilde{I}_m = \tilde{I}'_m, \quad \hat{I}_m = \hat{I}'_m, \qquad m = 1, 2, \ldots, \tag{2.5}$$

and

$$H(P) = H(P'), \qquad H(R) = H(R'), \tag{2.6}$$

then W and W' are spectrally equivalent. Moreover, for any linear function $F(\zeta)$, $\zeta \in C^r$,

$$(P \times P)(A(F)) = (P' \times P')(A'(F)), \tag{2.7}$$

$$(R \times R)(B(F)) = (R' \times R')(B'(F)), \tag{2.8}$$

$$(P \times R)(C(F)) = (P' \times R')(C'(F)), \tag{2.9}$$

where, e.g., $P \times R$ denotes the product measure on $\mathcal{X} \times \mathcal{Z}$ with marginals P and R.

COROLLARY. $\Pr\{(X(m), U) \in C(F)\} = \Pr\{(X'(m), U') \in C'(F)\}$ *for all F and m, and, in particular,*

$$\Pr\{(X, Z) \in C(F)\} = \Pr\{(X', Z') \in C'(F)\}, \qquad \text{all } F.$$

Having only used the invariants $\{I_m\}$, $\{\tilde{I}_m\}$, $\{\hat{I}_m\}$, $H(P)$, $H(R)$, we cannot hope for more than Theorem 2. Indeed (2.7)–(2.9) imply (2.5), and for W, W' non-singular, also (2.6).

The next two theorems say that, under some additional conditions on dist(X, Z), (2.6)–(2.9) imply full isomorphy of dist(X, Z) and dist(X', Z'). These additional conditions cannot hold if dist(X, Z) has non-trivial automorphisms, but, on the other hand, Theorem 4 holds for "almost all" joint distributions dist(X, Z), and Theorem 3 holds for "almost all" dist(X, Z) such that X and Z have equal ranges.

Denote by $\{C_1, \ldots, C_l\}$ the partition of $\mathcal{X} \times \mathcal{Z}$ into the non-void sets $C(F)$.

THEOREM 3. *Assume that* (i) W *and* $\tilde{W}$ *are non-singular*, (ii) *the eigenvalues* $\lambda_1, \ldots, \lambda_r$ *are log-independent*, (iii)

$$F_{xx} \neq F_{x\bar{x}} \quad \text{for } x \not\equiv \bar{x} \quad \text{and} \quad \tilde{F}_{zz} \not\equiv \tilde{F}_{z\bar{z}} \quad \text{for } z \neq \bar{z},$$

and (iv) *there is no partition* $\{D_1, \ldots, D_l\}$ *of* $\mathscr{X} \times \mathscr{Z}$, *different from* $\{C_1, \ldots, C_l\}$ *and such that*

$$(P \times R)(D_i) = (P \times R)(C_i), \qquad i = 1, \ldots, l.$$

If for some (X', Z') *the equalities* (2.5)–(2.6) *hold then* $\operatorname{dist}(X, Z)$ *and* $\operatorname{dist}(X', Z')$ *are isomorphic*.

If W has simple spectrum then denote by u_i the λ_i-eigenfunction of W, and

$$\tilde{u}_i = \frac{1}{\sqrt{\lambda_i}} V^* u_i.$$

Let $(\varepsilon_1, \ldots, \varepsilon_r)$ be a sequence of -1's and 1's. By $\operatorname{dist}(\varepsilon_1 u_1 \tilde{u}_1, \ldots, \varepsilon_r u_r \tilde{u}_r)$ we shall mean the joint distribution of the products $\varepsilon_i \cdot u_i \cdot \tilde{u}_i$, assuming

$$\operatorname{dist}(u_1, \ldots, u_r) = \operatorname{dist}(u_1(X), \ldots, u_r(X)),$$

$$\operatorname{dist}(\tilde{u}_1, \ldots, \tilde{u}_r) = \operatorname{dist}(\tilde{u}_1(Z), \ldots, \tilde{u}_r(Z)),$$

and $(u_1, \ldots, u_r)$ independent of $(\tilde{u}_1, \ldots, \tilde{u}_r)$.

THEOREM 4. *Let* W *be non-singular, and the eigenvalues* $\lambda_1, \ldots, \lambda_r$ *log-independent and simple. Assume further that for any* $(\varepsilon_1, \ldots, \varepsilon_r) \in \{-1, 1\}^r$, *the relation*

$$\operatorname{dist}(u_1 \tilde{u}_1, \ldots, u_r \tilde{u}_r) = \operatorname{dist}(\varepsilon_1 u_1 \tilde{u}_1, \ldots, \varepsilon_r u_r \tilde{u}_r) \tag{2.10}$$

implies at least one of the relations

$$\operatorname{dist}(u_1, \ldots, u_r) = \operatorname{dist}(\varepsilon_1 u_1, \ldots, \varepsilon_r u_r), \tag{2.11}$$

$$\operatorname{dist}(\tilde{u}_1, \ldots, \tilde{u}_r) = \operatorname{dist}(\varepsilon_1 \tilde{u}_1, \ldots, \varepsilon_r \tilde{u}_r). \tag{2.12}$$

If for some $\operatorname{dist}(X', Z')$ *the equalities* (2.5)–(2.6) *hold then* $\operatorname{dist}(X, Z)$ *and* $\operatorname{dist}(X', Z')$ *are isomorphic*.

COROLLARY. *Assume that* $\mathscr{X} = \mathscr{Z}$, $P = R$, V *is non-singular and self-adjoint with respect to* P,† *and the eigenvalues* $\lambda_1, \ldots, \lambda_r$ *are log-independent*

† I.e. $P(x)V(z \mid x) = P(z)V(x \mid z)$, $(x, z) \in \mathscr{X}^2$.

and simple. If for some dist(X', Z') *the equalities* (2.5)–(2.6) *hold then* dist(X, Z) *and* dist(X', Z') *are isomorphic.*

Although Theorems 3 and 4 may not be the strongest possible, full isomorphy of dist(X, Z), dist(X', Z') does not follow in general from (2.5)–(2.6). This will be demonstrated by joint distributions obtained from adjacency matrices of strongly regular graphs.

Let $\mathcal{X} = \mathcal{Z}$, P uniform on $\mathcal{X}$, and V a symmetric transition matrix on $\mathcal{X}^2$. Denote by $(Y(0), Y(1), \dots)$ a stationary Markov chain with dist$Y(0) = P$ and transition matrix V. Put

$$J_m = I(Y(0) \wedge Y(m)).$$

PROPOSITION (proved in the Appendix). *The sequence* $\{J_m\}$ *is not a full invariant for the isomorphy of symmetric joint distributions with a given alphabet size and uniform marginals.*

§3. Proof of Theorem 1

PROOF OF LEMMA 1. For z and m fixed, denote

$$f_z(x) = \frac{V^{*(m)}(x \mid z)}{P(x)} - 1.$$

Expand f_z in the system $\{u_{ij}\}_{i=1,\dots,r, j=1,\dots,s_i} \cup \{u_{0j}\}$ (where $\{u_{0j}\}$ denotes a complete orthonormal set of 0-eigenfunctions of W):

$$f_z(x) = \sum_{i,j} \gamma_z(i,j) u_{ij}(x)$$

with

$$\begin{aligned}\gamma_z(i,j) &= \sum_x P(x) f_z(x) u_{ij}(x) \\ &= \sum_x V^{*(m)}(x \mid z) u_{ij}(x) \\ &= \lambda_i^{m+1/2} \tilde{u}_{ij}(z),\end{aligned}$$

whence

$$W^{*(m)}(x \mid z) = P(x) \hat{F}_{xz}(\lambda^{m+1/2}),$$

proving (i). Applying (i) for $m = 0$ and with W^m in the role of V^*, we get (ii).

The first equality in (iii) follows from (i) and the definition of mutual information. To obtain the second equality, use the power series of the function $(1+\vartheta)\log(1+\vartheta)$, $\vartheta \in \mathbf{C}$, around $\vartheta = 0$. (iv) is proved similarly. □

Denote $\pi_1, \pi_2, \ldots$ the different products $\lambda_1^{i_1} \cdot \cdots \cdot \lambda_r^{i_r}$; then the expansion (2.2) can be written as

$$I_m = \sum_k c_k \pi_k^m. \tag{3.1}$$

LEMMA 2. (i) *For $\lambda_j \in \Lambda$, the coefficient of λ_j^{2m} in* (3.1) *is half the multiplicity of λ_j.*

(ii) *For $\lambda_j \notin \Lambda$,*

$$\lambda_j = \prod_{\lambda_i \in \Lambda} \lambda_i^{\alpha_{ji}}, \quad \alpha_{ji} \geqq 0 \text{ rationals}, \quad \sum_i \alpha_{ji} \geqq \tfrac{3}{2}. \tag{3.2}$$

PROOF. (i) follows from the fact that the coefficients of the terms λ_i^m and $\lambda_i^m \lambda_j^m$ $(i \neq j)$ in (2.2) are 0.

To prove (ii), assume $\Lambda = \{\lambda_1, \ldots, \lambda_l\}$. For $j \geqq l+1$ we have

$$\lambda_j = \prod_i \lambda_i^{\alpha_{ji}}, \quad \alpha_{ji} \geqq 0 \text{ rationals}, \quad \sum_i \alpha_{ji} \geqq \tfrac{3}{2}. \tag{3.3}$$

It is clear that we must have $\alpha_{jj} < 1$, so (3.3) can be brought to the form

$$\lambda_j = \prod_i \lambda_i^{\beta_{ji}}, \quad \beta_{ji} \geqq 0 \text{ rationals}, \quad \beta_{jj} = 0, \quad \sum_i \beta_{ji} \geqq \tfrac{3}{2}. \tag{3.4}$$

Substituting (3.4), taken with $j = r$, into (3.4), taken with $j = r-1$, we get

$$\lambda_{r-1} = \lambda_r^{\beta_{r-1,r}} \prod_{i \leqq r-2} \lambda_i^{\beta_{r-1,r} \cdot \beta_{r,i} + \beta_{r-1,i}}. \tag{3.5}$$

(3.4) implies

$$\beta_{r-1,r} \cdot \beta_{r,r-1} + \sum_{i \leqq r-2} (\beta_{r-1,r} \cdot \beta_{r,i} + \beta_{r-1,i})$$

$$= \beta_{r-1,r} \sum_i \beta_{r,i} + \sum_{i \leqq r-2} \beta_{r-1,i} \geqq \sum_i \beta_{r-1,i} \geqq \tfrac{3}{2}.$$

Since $\beta_{r-1,r} \cdot \beta_{r,r-1}$ must be < 1, (3.5) can be brought to the form

$$\lambda_{r-1} = \prod_{i \leqq r-1} \lambda_i^{\gamma_{r-1,i}}, \quad \gamma_{r-1,i} \geqq 0 \text{ rationals}, \quad \gamma_{r-1,r-1} = 0, \quad \sum_i \gamma_{r-1,i} \geqq \tfrac{3}{2}.$$

Similarly, for any $l+1 \leqq j \leqq r-1$,

$$\lambda_j = \prod_{i \leqq r-1} \lambda_i^{\gamma_{ji}}, \quad \gamma_{ji} \geqq 0 \text{ rationals}, \qquad \gamma_{ji} = 0, \quad \sum_i \gamma_{ji} \geqq \tfrac{3}{2}.$$

Iterating this step, we get the expressions (3.2) for $j \geqq l+1$. □

PROOF OF THEOREM 1. Assume $I_m = I'_m$ for some dist(X', Z'). Then the series (3.1) for dist(X, Z) and dist(X', Z') coincide. Denote $\mu_1, \ldots, \mu_t$ the different eigenvalues of W' in $(0, 1)$.

We claim that $\lambda_j \in \Lambda \setminus \Lambda'$ would imply

$$\lambda_j = \prod_i \mu_i^{\gamma_{ji}}, \quad \gamma_{ji} \geqq 0 \text{ rationals}, \quad \sum_i \gamma_{ji} \geqq \tfrac{3}{2}. \tag{3.6}$$

Indeed, for $\lambda_j \in \{\mu_1, \ldots, \mu_t\} \setminus \Lambda'$ this follows from the definition of Λ'. For $\lambda_j \in \Lambda \setminus \{\mu_1, \ldots, \mu_t\}$, the coefficient of λ_j^{2m} in (2.2) is $\neq 0$, so λ_j^2 must be of the form $\Pi_i \mu_i^{\beta_{ji}}$, $\beta_{ji} \geqq 0$ integers, and since $\lambda_j \notin \{\mu_1, \ldots, \mu_t\}$, we must have $\Sigma_i \beta_{ji} \geqq 3$ which implies (3.6).

On the other hand, it easily follows from the previous statement, and Lemma 2, that μ_i $(i = 1, \ldots, t)$ can be written as

$$\mu_i = \prod_k \lambda_k^{\delta_{ik}}, \quad \delta_{ik} \geqq 0 \text{ rationals}, \quad \sum_k \delta_{ik} \geqq 1. \tag{3.7}$$

Substituting (3.7) into (3.6), we get

$$\lambda_j = \prod_k \lambda_k^{\sum_i \gamma_{ji}\delta_{ik}}.$$

Since $\lambda_j \in \Lambda$, we must have $\Sigma_{i,k} \gamma_{ji}\delta_{ik} < \frac{3}{2}$. But $\Sigma_{i,k} \gamma_{ji}\delta_{ik} \geqq \Sigma_i \gamma_{ji} \geqq \frac{3}{2}$, a contradiction proving $\Lambda = \Lambda'$. The statement on the multiplicities is obvious. □

§4. Proof of Theorem 2

We shall need the following lemmas.

LEMMA 3. *Let $\varphi(\zeta) = \varphi(\zeta_1, \ldots, \zeta_r)$ be a complex function analytic in a region $\Sigma_1^r |\zeta_i| < \delta$. Let the numbers $0 < \lambda_1, \lambda_2, \ldots, \lambda_r < 1$ be log-independent. If*

$$\varphi(\lambda_1^n, \ldots, \lambda_r^n) = 0$$

for n large enough then $\varphi(\zeta) = 0$ in $\Sigma_1^r |\zeta_i| < \delta$.

PROOF. We may assume $\Sigma \lambda_i \prec \delta$. Consider the power series of φ around the origin

$$\varphi(\zeta) = \sum_{l_1,\ldots,l_r=0}^{\infty} a(l_1,\ldots,l_r)\zeta_1^{l_1}\cdots\zeta_r^{l_r}. \tag{4.1}$$

We have

$$0 = \varphi(\lambda_1^n,\ldots,\lambda_r^n) = \sum_{l_1,\ldots,l_r} a(l_1,\ldots,l_r)\lambda_1^{l_1 n}\cdots\lambda_r^{l_r n} \tag{4.2}$$

for n large enough. By log-independence, the products $\lambda_1^{l_1}\cdots\lambda_r^{l_r}$ are all distinct. Denote $v_1, v_2, \ldots$ these products in decreasing order, and put $b_i = a(l_1,\ldots,l_r)$ if $v_i = \lambda_1^{l_1}\cdots\lambda_r^{l_r}$. Since the series (4.1) is absolutely convergent in $\Sigma\,|\zeta_i| < \delta$, (4.2) may be written as

$$\sum_i b_i v_i^n = 0 \qquad \text{for } n \text{ large enough.} \tag{4.3}$$

It is enough to prove $b_i = 0$ for all i.

Assume $b_0 = b_1 = \cdots = b_{q-1} = 0$, $b_q \neq 0$ for some $q \geqq 0$. Then $b_p \neq 0$ for some $p > q$; let p denote the smallest such integer. By (4.3),

$$b_q = \left(\frac{v_p}{v_q}\right)^n \left[b_p + b_{p+1}\left(\frac{v_{p+1}}{v_p}\right)^n + b_{p+2}\left(\frac{v_{p+2}}{v_p}\right)^n + \cdots \right],$$

whence

$$|b_q| \leqq \left(\frac{v_p}{v_q}\right)^n \sum_{j=0} |b_{p+j}| \left(\frac{v_{p+j}}{v_p}\right)^n \leqq \left(\frac{v_p}{v_q}\right)^n \frac{1}{v_p} \sum_{j \geqq 0} |b_{p+j}|\, v_{p+j}.$$

Since $v_p/v_q < 1$ and $\Sigma_j\, |b_{p+j}|\, v_{p+j} < \infty$, this implies $b_q = 0$. □

LEMMA 4. *W is non-singular iff for any $x, y \in \mathscr{X}^2$*

$$1 + \sum_{i=1}^{r} \sum_{j=1}^{s_i} u_{ij}(x)u_{ij}(y) = \frac{\delta(xy)}{P(x)} \qquad \text{with } \delta(x,y) = \begin{cases} 0, & x \neq y, \\ 1, & x = y. \end{cases} \tag{4.4}$$

PROOF. Let $\{u_{0j}\}_{j=1,\ldots,s_0}$ denote a (possibly void) complete orthonormal system of 0-eigenfunctions of W. Then the functions $\equiv 1$ and $\{u_{ij}(x)\}_{i=0,\ldots,r\ j=1,\ldots,s_i}$ constitute a complete orthonormal system with respect to P. It is easy to see that this implies

$$1 + \sum_{i=0}^{r} \sum_{j=1}^{s_i} u_{ij}(x)u_{ij}(y) = \frac{\delta(xy)}{P(x)}, \qquad \text{all } x, y.$$

Thus (4.4) is equivalent to

$$\sum_{j=1}^{s_0} u_{0j}(x)u_{0j}(y) = 0, \qquad \text{all } x, y. \qquad \square$$

COROLLARY. *If W is non-singular then $F_{xy} \not\equiv 1$.*

PROOF. $\delta(x, y)/P(x) \neq 1$ for any x, y, so $F_{xy}(1, \ldots, 1) \neq 1$. $\square$

PROOF OF THEOREM 2. Denote $\mu_1, \ldots, \mu_t$ the different eigenvalues of W' in $(0, 1)$. By Theorem 1, we may assume $\mu_i = \lambda_i$, $i \leqq r$. By Lemma 2, there exist a natural s and integers a_{ji}, $r < j \leqq t$, $1 \leqq i \leqq r$, such that for $\nu_i = \lambda_i^{1/s}$,

$$\mu_j = \prod_{i=1}^{r} \nu_i^{a_{ji}}, \quad \sum_i a_{ji} > s, \qquad j = r+1, r+2, \ldots, t. \tag{4.5}$$

Denote

$$G_{xy}(\zeta_1, \ldots, \zeta_r) = F_{xy}(\zeta_1^s, \ldots, \zeta_r^s),$$

$$G'_{x'y'}(\zeta_1, \ldots, \zeta_r) = F'_{x'y'}\left(\zeta_1^s, \ldots, \zeta_r^s, \prod_1^r \zeta_i^{a_{r+1,i}}, \ldots, \prod_1^r \zeta_i^{a_{ti}}\right),$$

$$\Phi(\zeta) = \sum_{x,y} P(x)P(y)G_{xy}(\zeta)\log G_{xy}(\zeta), \tag{4.6}$$

$$\Phi'(\zeta) = \sum_{x',y'} P'(x')P'(y')G'_{x'y'}(\zeta)\log G'_{x'y'}(\zeta) \qquad (\zeta \in \mathbf{C}^r),$$

$$\Psi(\bar{\zeta}) = \Phi(\bar{\zeta}, \ldots, \bar{\zeta}), \quad \Psi'(\bar{\zeta}) = \Psi'(\bar{\zeta}, \ldots, \bar{\zeta}), \qquad \bar{\zeta} \in \mathbf{C}.$$

The functions Φ, Φ' are analytical in a neighborhood of 0, and so are Ψ and Ψ'. Ψ and Ψ' can be continued to functions analytical in a region obtained by removing finitely many half-lines from the complex plane.

We have

$$I_m = \Phi(\nu^m), \qquad I'_m = \Phi'(\nu^m),$$

so by Lemma 3, $\Phi(\zeta) = \Phi'(\zeta)$, and, *a fortiori*,

$$\Psi(\bar{\zeta}) = \Psi'(\bar{\zeta}) \quad \text{in a neighborhood of 0.} \tag{4.7}$$

In order to prove that W' is non-singular, let us show first that $\Psi'(\bar{\zeta})$ is analytical on the segment $0 \leqq \zeta < 1$. Indeed, W being non-singular, Lemma 4 implies (4.4). Hence $G_{xy}(\bar{\zeta}, \ldots, \bar{\zeta}) > 0$ for $0 \leqq \bar{\zeta} < 1$, which implies the analyticity of Ψ, and, consequently, of Ψ', for $0 \leqq \bar{\zeta} < 1$. This, in turn, implies

$G'_{x'y'}(\zeta, \ldots, \zeta) > 0$ for $(x', y') \in \mathscr{X}'^2$ and $0 \leqq \zeta < 1$, whence $G'_{x'y'}(1, \ldots, 1) \geqq 0$ for all (x', y').

Consider the probability distributions

$$Q(x, y) = P(x)P(y)G_{xy}(1, \ldots, 1), \qquad (x, y) \in \mathscr{X}^2,$$

$$Q'(x', y') = P'(x')P'(y')G'_{x'y'}(1, \ldots, 1), \qquad (x', y') \in \mathscr{X}'^2.$$

By (4.4)

$$Q(x, y) = P(x)\delta(x, y),$$

which implies

$$\Psi(1) = H(P),$$

whence

$$\Psi'(1) = H(P'). \tag{4.8}$$

But both marginals of Q' coincide with P', so (4.8) implies

$$Q'(x', y') = \delta(x', y')P'(x'),$$

i.e.

$$1 + \sum_{i,j} u'_{ij}(x')u'_{ij}(y') = \frac{\delta(x', y')}{P'(x')},$$

which is equivalent to the non-singularity of W', by Lemma 4.

Now, Corollary to Lemma 4 implies that the degree of $G'_{x'y'}$ is $\geqq s$ for all (x', y').† Comparing the singularities of the functions Ψ, Ψ', we get from (4.7)

$$\begin{aligned} s &= \sum_{x,y} P(x)P(x)\deg[G_{xy}(\zeta, \ldots, \zeta)] \\ &= \sum_{x',y'} P'(x')P'(y')\deg[G'_{x'y'}(\zeta, \ldots, \zeta)] \\ &\geqq s, \end{aligned}$$

whence $\deg G'_{x'y'} = s$ for all x', y'. By (4.4) and (4.5), this implies that

$$\{\mu_1, \ldots, \mu_t\} = \{\lambda_1, \ldots, \lambda_r\},$$

and by Theorem 1, W and W' are spectrally equivalent.

† In Thouvenot's proof of Proposition 1 [10], it is not proved that $P_{kl}(u)$ is not a constant, and without this, the statement of the Proposition does not follow from (5).

Now $I_m = I'_m$ implies, using Lemma 3,

$$\sum_{xy} P(x)P(y)F_{xy}(\zeta)\log F_{xy}(\zeta) = \sum_{x'x'} P'(x')P(y')F'_{x'y'}(\zeta)\log F'_{x'y'}(\zeta),$$

and since the polynomials F_{xy}, $F'_{x'y'}$ are linear, (2.7) follows easily.

Since W and W' are spectrally equivalent, the positive eigenvalues of $\tilde{W}$ and $\tilde{W}'$, and their multiplicities, coincide. ($\tilde{W}$ and/or $\tilde{W}'$ may be singular.) Denote

$$\tilde{\phi}(\zeta) = \sum_{z,t} R(z)R(t)\tilde{F}_{zt}(\zeta)\log \tilde{F}_{zt}(\zeta),$$

and define $\tilde{\phi}'$ similarly. Then, by Lemma 3,

$$\tilde{\phi}(\zeta) = \tilde{\phi}'(\zeta),$$

and since F_{zt} and $F'_{z't'}$ are linear, this implies (2.8). (2.9) can be proved similarly. The Corollary follows from (2.9) and (i) of Lemma 1. □

§5. Proof of Theorems 3 and 4

PROOF OF THEOREM 3. By (i) and (ii), Theorem 2 applies. By (iii) there exist bijections $\sigma: \mathscr{X} \to \mathscr{X}'$, $\tau: \mathscr{Z} \to \mathscr{Z}'$ satisfying

$$P(x) = P'(\sigma(x)), \qquad R(z) = R'(\tau(z)),$$

and

$$F_{xx} = F'_{\sigma(x),\sigma(x)}, \qquad \tilde{F}_{zz} = \tilde{F}'_{\tau(z),\tau(z)}.$$

By (iv), (σ, τ) is an isomorphism. □

To prove Theorem 4, we need

LEMMA 5. *Let $\mathscr{X}_1, \dots, \mathscr{X}_r$ be finite sets of real numbers, and q, q' real valued functions on $\mathscr{X}_1 \times \cdots \times \mathscr{X}_r$. If*

$$\left| \sum_{\substack{(x_1,\dots,x_r) \\ \in \mathscr{X}_1 \times \cdots \times \mathscr{X}_r}} q(x_1, \dots, x_r)x_1^{i_1} \cdots x_r^{i_r} \right| = \left| \sum_{\substack{(x_1,\dots,x_r) \\ \in \mathscr{X}_1 \times \cdots \times \mathscr{X}_r}} q'(x_1, \dots, x_r)x_1^{i_1} \cdots x_r^{i_r} \right|$$

for any r-tuple of integers $(i_1, \dots, i_r) \geqq 0$ then there exists a sequence $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_r) \in \{-1, 1\}^{r+1}$ such that for all $x_1, \dots, x_r$,

$$q(x_1, \dots, x_r) = \varepsilon_0 q'(\varepsilon_1 x_1, \dots, \varepsilon_r x_r).$$

PROOF. Let us prove the statement for $r = 1$ first. For any i,

$$\sum_{x>0}(q(x)+q(-x))x^{2i}=\delta_0(i)\sum_{x>0}(q'(x)+q'(-x))x^{2i},$$

$$\sum_{x>0}(q(x)-q(-x))x^{2i+1}=\delta_1(i)\sum_{x>0}(q'(x)-q'(-x))x^{2i+1},$$

where $|\delta_0(i)|=|\delta_1(i)|=1$. There exists a fixed assignment of δ_0 and δ_1 such that for infinitely many i, $\delta_0(i)=\delta_0$ and $\delta_1(i)=\delta_1$. For this δ_0 and δ_1,

$$q(x)+q(-x)=\delta_0(q'(x)+q'(-x)),$$

$$q(x)-q(-x)=\delta_1(q'(x)-q'(-x)),$$

for any $x>0$.

It follows that for $x>0$

$$q(x)=\tfrac{1}{2}(\delta_0+\delta_1)q'(x)+\tfrac{1}{2}(\delta_0-\delta_1)q'(-x),$$

$$q(-x)=\tfrac{1}{2}(\delta_0+\delta_1)q'(-x)+\tfrac{1}{2}(\delta_0-\delta_1)q'(x),$$

which implies the desired result for $r=1$. For $r>1$, it follows easily by induction. □

PROOF OF THEOREM 4. By Theorem 2, W and W' are spectrally equivalent, and both have simple spectrum. By statement (iv) of Lemma 1, this implies

$$\left|\sum_x P(x)u_1^{i_1}(x)\cdots u_r^{i_r}(x)\right|=\left|\sum_{x'}P'(x')u_1'^{i_1}(x')\cdots u_1'^{i_r}(x')\right|$$

for any r-tuple of non-negative integers. Since P and P' are non-negative, Lemma 5 implies that, after possibly multiplying some of the functions u_i' by -1, we shall have

$$\operatorname{dist}(u_1,\ldots,u_r)=\operatorname{dist}(u_1',\ldots,u_r'). \tag{5.1}$$

Similarly,

$$\operatorname{dist}(\tilde{u}_1,\ldots,\tilde{u}_r)=\operatorname{dist}(\varepsilon_1\tilde{u}_1,\ldots,\varepsilon_1\tilde{u}_r) \tag{5.2}$$

for some $(\varepsilon_1,\ldots,\varepsilon_r)\in\{-1,1\}^r$. Moreover, Theorem 2 implies

$$\operatorname{dist}(u_1\tilde{u}_1,\ldots,u_r\tilde{u}_r)=\operatorname{dist}(u_1'\tilde{u}_1',\ldots,u_r'\tilde{u}_r'). \tag{5.3}$$

(5.1), (5.2) and (5.3) imply (2.10), and hence either (2.11) or (2.12). If (2.12) holds then (5.2) may be replaced by

$$\operatorname{dist}(\tilde{u}_1, \ldots, \tilde{u}_r) = \operatorname{dist}(\tilde{u}'_1, \ldots, \tilde{u}'_r).$$

Since (X, Z) and (X', Z') are non-singularly dependent, $(u_1, \ldots, u_r)$, $(\tilde{u}_1, \ldots, \tilde{u}_r)$, $(u'_1, \ldots, u'_r)$, $(\tilde{u}'_1, \ldots, \tilde{u}'_r)$ separate $\mathscr{X}$, $\mathscr{Z}$, $\mathscr{X}'$ and $\mathscr{Z}'$, respectively (cf. [6]). Now the statement follows from (i) of Lemma 1. The case when (2.11) holds can be settled similarly.

The Corollary follows from the fact that for V self-adjoint, $u_i \equiv \pm \tilde{u}_i$, so (2.10) implies (2.11). □

Appendix

PROOF OF LEMMA 0. Consider the function†

$$T_{P,V}(t) = \min\{H(Z \mid S) : \operatorname{dist}(S, X, Z) \text{ satisfies } I(S \wedge Z \mid X) = 0,$$
$$\operatorname{dist} X = P, \operatorname{dist}(Z \mid X) = V,$$
$$H(X \mid S) \geqq t, |S| \leqq |X| + 2\}$$

$(0 \leqq t \leqq H(X))$. It follows from the results in [1] or [11] that this function is invariant for weak asymptotic isomorphy, as pointed out by Gács and Körner [5].

On the other hand, it is known ([3], Problem 28 of §4, Chapter 3) that 1 is a multiple eigenvalue of W if and only if

$$T_{P,V}(t) = H(Z) - H(X) + t \quad \text{for } t_1 \leqq t \leqq H(X) \qquad (t_1 < H(X)).$$

Moreover, the fact that V has at least two identical rows is equivalent to the existence of a function $S = \varphi(X)$ satisfying $H(S) > 0$ and $I(X \wedge Z \mid S) = 0$ which is obviously equivalent to:

$$T_{P,V}(t) = H(Z \mid X) \quad \text{for } 0 \leqq t \leqq t_0 \qquad (t_0 > 0),$$

as can be easily seen. □

To prove the Proposition, we construct symmetric transition matrices from adjacency matrices of strongly regular graphs.

Let G be an undirected graph with vertex set $\mathscr{X}$, without loops and multiple edges. G is called regular with degree r if every vertex of G is adjacent to r vertices.

† $|S|$ denotes the size of the range of the random variable S. $I(S \wedge Z \mid X)$ denotes average conditional mutual information. $H(X \mid S)$ denotes average conditional entropy.

DEFINITION (see [4]). G is strongly regular with parameters t, r, e, f ($e, f \geqq 0$ integers) if $|\mathscr{X}| = t$, G is regular with degree r, and its adjacency-matrix A satisfies

$$A^2 = (e - f)A + fJ + (r - f)I, \tag{A.1}$$

where I is the identity matrix, and J is the matrix with all elements equal to 1.

PROOF OF THE PROPOSITION. Let P be uniform on $\mathscr{X}$, G a strongly regular graph on $\mathscr{X}$, and

$$V(z \mid x) = \frac{1}{r} A(x, z).$$

V is a symmetric transition matrix. From (A.1) it follows by induction that

$$W^m = \alpha_m A + \beta_m J + \gamma_m I,$$

where the constants α_m, β_m, γ_m are determined by the parameters t, r, e, f. Hence $H(Y(m) \mid Y(0))$ is determined by the numbers t, r, e, f, too. Since there exist non-isomorphic strongly regular graphs with the same parameters (see [2]), this proves the Proposition. □

REFERENCES

1. R. Ahlswede and J. Körner, *Source coding with side information at the decoder and a converse for degraded broadcast channels*, IEEE Trans. Inf. Theory **21** (1975), 629–637.

2. F. C. Bussemaker and J. J. Seidel, *Symmetric Hadamard matrices of order 36*, Ann. N.Y. Acad. Sci. **175** (1970), 66–79.

3. I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, Akadémiai Kiadó, Budapest, and Academic Press, New York, 1981.

4. D. M. Cvetkovič, M. Doob and H. Sachs, *Spectra of Graphs. Theory and Application*, Academic Press, New York, 1979.

5. P. Gács and J. Körner, oral communication.

6. K. Marton, *The problem of isomorphy for general discrete memoryless sources*, Z. Wahrscheinlichkeitstheor. Verw. Geb. **63** (1983), 51–58.

7. K. Marton, *Sequences achieving the boundary of the entropy region for a 2-source are virtually memoryless*, IEEE Trans. Inf. Theory, to appear.

8. K. Marton, *On the problem of asymptotic isomorphy for discrete memoryless stationary correlated sources*, Z. Wahrscheinlichkeitstheor. Verw. Geb., submitted.

9. M. S. Pinsker, *Information and Information Stability of Random Variables and Processes* (trans. A. Feinstein), Holden-Day, San Francisco, 1964.

10. J.-P. Thouvenot, *Remarques sur des systèmes dynamiques donnés avec plusieurs facteurs*, Isr. J. Math. **21** (1975), 215–232.

11. A. D. Wyner, *On source coding with side information at the decoder*, IEEE Inf. Theory **21** (1975), 294–300.