Contemporary Mathematics
Volume 149, 1993

Majorization, Monotonicity of Relative Entropy, and Stochastic Matrices

JOEL E. COHEN, YVES DERRIENNIC AND GH. ZBAGANU

ABSTRACT. Schur showed that $x = Ay$ implies $\sum_i g(x_i) \le \sum_i g(y_i)$ for any positive probability n-vectors x and y, any doubly stochastic $n \times n$ matrix A, and any convex function $g:(0,1) \to \Re$. We establish a quantitative improvement of Schur's theorem: under the same hypotheses,

$$\sum_i g(x_i) \le \bar{\alpha}(A) \sum_i g(y_i) + \alpha(A) n g(1/n) \le \sum_i g(y_i),$$

where

$$\bar{\alpha}(A) \equiv (1/2)\max_{j,k} \textstyle\sum_{i=1}^{m} |a_{ij} - a_{ik}| \text{ and } \alpha(A) = 1 - \bar{\alpha}(A).$$

This improvement follows from a recent quantitative sharpening of the monotonicity theorem of relative entropy. We also establish a converse of the monotonicity theorem of relative entropy (sometimes called the Data Processing Lemma). Specifically, for any positive probability n-vectors x and y and any positive probability m-vectors u and v, if $H_\phi(u,v) \le H_\phi(x,y)$ for every relative ϕ-entropy H_ϕ, then there exists a row-allowable column-stochastic $m \times n$ matrix A such that $u = Ax$ and $v = Ay$.

1. Introduction

Let m and n be finite positive integers. Let P_n be the set of positive probability column-vectors with n elements, i.e., $P_n \equiv \{x \in \Re^n : x_i > 0 \; \forall i, \; \sum_i x_i = 1\}$. An $n \times n$ matrix is doubly stochastic if its elements are nonnegative real numbers, every row has sum 1 and every column has sum 1. As usual, a real-valued function h on some convex subset D of a vector space over the reals is called convex if, for all $p \in [0, 1]$ and all $s, t \in D$, $h(ps + [1-p]t) \le ph(s) + (1-p)h(t)$. Recall that a convex function on a convex open subset U of $\Re^n$ is continuous on U (e.g., Roberts and Varberg 1973, p. 93). A fundamental inequality of the theory of majorization (e.g., Marshall and Olkin 1979, p. 108) states:

1991 Mathematics Subject Classification. Primary 15A51; Secondary 15A90, 60J10.
Supported in part by U.S. National Science Foundation grant BSR 87-05047.
This paper is in final form and no version of it will be submitted for publication elsewhere.

THEOREM 1.1. *Let $x, y \in P_n$. The inequality $\sum_i g(x_i) \le \sum_i g(y_i)$ holds for all convex functions $g:(0,1) \to \Re$ if and only if there exists a doubly stochastic $n \times n$ matrix A such that $x = Ay$.*

Marshall and Olkin (1979) attribute the equivalent of this result to Hardy, Littlewood and Pólya (1929) and Karamata (1932), and the sufficiency portion (i.e., $x = Ay$ implies $\sum_i g(x_i) \le \sum_i g(y_i)$) to Schur (1923). See also Hardy, Littlewood and Pólya (1952, ¶3.17) and Alberti and Uhlmann (1982). Csiszár and Körner (1981, p. 58, Exercise 14) use the monotonicity of relative entropy (which they call the Data Processing Lemma (their Lemma 3.11, p. 55)) to establish the sufficiency part of Theorem 1.1 (Schur's theorem) in the special case where $g(s) = s \log s$.

The first purpose of this note is to sharpen the inequality $\sum_i g(x_i) \le \sum_i g(y_i)$ in Schur's theorem (see Theorem 2.1). The improvement follows from a recent quantitative sharpening of the monotonicity theorem of relative entropy (see Theorem 1.4). An open question is whether, conversely, Schur's theorem (in either its original form or as sharpened in Theorem 2.1) implies the monotonicity of relative entropy (in its original form or as sharpened in Theorem 1.4).

The second purpose of this note is to establish (in Theorem 4.1) a converse of the monotonicity theorem of relative entropy. We shall prove elsewhere that this converse holds in a sharpened, quantitative form.

To state results, some definitions are required. A real-valued function h on some convex cone D of a vector space over the reals is called homogeneous (meaning homogeneous of degree one) if, for all $x \in D$ and all nonnegative λ, $h(\lambda x) = \lambda h(x)$.

DEFINITION 1.2. Let ϕ be a real-valued function on $(0, \infty) \times (0, \infty)$ that is homogeneous and jointly convex in its arguments and satisfies $\phi(1, 1) = 0$. For any two positive n-vectors $x = (x_i)$ and $y = (y_i)$, whether or not x and y are probability vectors, define the relative ϕ-entropy $H_\phi(x, y)$ by $H_\phi(x, y) = \sum_i \phi(x_i, y_i)$.

This generalization of relative entropy has been widely studied under various names and notations (e.g., Liese and Vajda 1987). Any real-valued function g that is convex on $(-1, \infty)$ with $g(0) = 0$ can be used to define ϕ that satisfies Definition 1.2 by putting $\phi(x, y) = xg((y/x) - 1)$. Thus, as examples,

$$g(t) = |t| \Rightarrow H_\phi(x, y) = \sum_i |x_i - y_i|,$$

$$g(t) = t^2 \Rightarrow H_\phi(x, y) = \sum_i \frac{(x_i - y_i)^2}{x_i},$$

$$g(t) = -\log(1+t) \Rightarrow H_\phi(x, y) = \sum_i x_i \log \frac{x_i}{y_i},$$

$$g(t) = (1+t)\log(1+t) \Rightarrow H_\phi(x, y) = \sum_i y_i \log \frac{y_i}{x_i},$$

$$g(t) = t \log(1+t) \Rightarrow H_\phi(x, y) = \sum_i (y_i - x_i) \log \frac{y_i}{x_i}.$$

Except possibly for constants, the first expression on the right is the l_1 norm, the second is the Pearson χ^2-statistic for goodness of fit, the third is the Kullback-Leibler divergence or relative entropy of information theory (Csiszár and Körner 1981) or the G^2 likelihood ratio statistic in the theory of contingency tables, the fourth is the same with the roles of x and y exchanged, and the last (which is the sum of the preceding two) is the entropy production of statistical physics or the symmetric divergence of information theory. Thus significantly diverse measures are subsumed under the generalization of ϕ-entropy.

As usual, the l_p-norms are defined for a vector x and for $1 \le p < \infty$ by $\|x\|_p = (\sum |x_i|^p)^{1/p}$. A column-stochastic matrix is an $m \times n$ matrix with each element a nonnegative real number and with all column sums 1.

DEFINITION 1.3. For any column-stochastic $m \times n$ matrix A, Dobrushin's (1956) coefficient of ergodicity is

$$\alpha(A) = \min_{j,k} \sum_{i=1}^{m} \min(a_{ij}, a_{ik}) .$$

The complement $1 - \alpha(A)$ will be written

$$\bar{\alpha}(A) \equiv 1 - \alpha(A) = \frac{1}{2} \max_{j,k} \sum_{i=1}^{m} | a_{ij} - a_{ik} |$$

and satisfies (Dobrushin 1956, pp. 69-70)

$$\bar{\alpha}(A) = \sup\left\{ \frac{\| A(x-y) \|_1}{\| x-y \|_1} : x \text{ and } y \text{ are positive } n\text{-vectors, } x \neq y,\ \| x \|_1 = \| y \|_1 \right\}.$$

A matrix is row-allowable if each row contains at least one positive element. Every doubly stochastic matrix is clearly stochastic and row-allowable. A column-stochastic $m \times n$ matrix is called a scrambling matrix (Hajnal 1958, p. 235) if any submatrix consisting of two columns has a row both elements of which are positive; i.e., $A = (a_{ij})$ is scrambling if, for all j and k such that $1 \le j < k \le n$, there exists an i such that $1 \le i \le m$ and $a_{ij}a_{ik} > 0$. A column-stochastic, row-allowable matrix A is scrambling if and only if $\alpha(A) > 0$.

THEOREM 1.4. *Let A be a column-stochastic, row-allowable $m \times n$ matrix and let $x, y \in P_n$. Then*

$$H_\phi(Ax, Ay) \le \bar{\alpha}(A) H_\phi(x, y) .$$

The significant feature of this theorem, due to Cohen, Iwasa, Rautu, Ruskai, Seneta and Zbaganu (in press), is that the coefficient $\bar{\alpha}(A)$ is valid regardless of which ϕ-entropy is chosen and for all $x, y \in P_n$. $\bar{\alpha}(A) < 1$ if and only if A is a scrambling matrix.

W. Doeblin developed an early quantitative measure of the contractive action of a stochastic matrix. See Seneta (1973) for a historical perspective.

DEFINITION 1.5. If A is a column-stochastic, row-allowable $m \times n$ matrix, Doeblin's (1937) coefficient of ergodicity δ is

$$\delta(A) = \sum_{i=1}^{m} \min\{a_{ij} : j = 1, \ldots, n\} .$$

Define $\bar{\delta}(A) = 1 - \delta(A)$.

It is known that $\delta(A) \le \alpha(A)$ with equality if $n=2$. Thus $\overline{\delta}(A) \ge \overline{\alpha}(A)$ (with equality if $n=2$). Therefore the inequalities in Theorem 1.4 and Theorem 2.1 hold under the same hypotheses with α replaced by δ and $\overline{\alpha}$ replaced by $\overline{\delta}$.

2. Quantitative majorization in discrete processes

THEOREM 2.1. *Let $h : \mathfrak{R} \to \mathfrak{R}$ be a convex function, A a doubly stochastic $n \times n$ matrix, $x \in P_n$ (i.e., x is a strictly positive probability vector), and $w = Ax$. Then*

$$\sum_i h(w_i) \le \overline{\alpha}(A) \sum_i h(x_i) + \alpha(A) n h\left(\frac{1}{n}\right) \le \sum_i h(x_i).$$

Proof. Define the function $\phi : \mathfrak{R} \times \mathfrak{R} \to \mathfrak{R}$ by the requirements that ϕ be homogeneous and that

$$\phi(1,\ 1+s) = h\left(\frac{s+1}{n}\right) - h\left(\frac{1}{n}\right), \qquad \forall s.$$

Then ϕ is jointly convex in both arguments and $\phi(1,\ 1) = 0$. Therefore ϕ satisfies the hypotheses of Theorem 1.4.

The doubly stochastic matrix A is necessarily column-stochastic and row-allowable and therefore satisfies the hypotheses of Theorem 1.4. Choose y to be the n-vector with each element $1/n$. Then $Ay = y$. Let $\overline{\alpha} = \overline{\alpha}(A)$, $\alpha = 1 - \overline{\alpha}$. Theorem 1.4 implies

$$H_\phi(y,\ w) \le \overline{\alpha} H_\phi(y,\ x).$$

But

$$H_\phi(y,\ x) = \sum_i \phi\left(\frac{1}{n},\ x_i\right) = \sum_i \phi\left(\frac{1}{n},\ \frac{1}{n} n x_i\right) = \frac{1}{n} \sum_i \phi(1,\ (nx_i - 1) + 1)$$

$$= \frac{1}{n} \sum_i \left[h\left(\frac{(nx_i - 1) + 1}{n}\right) - h\left(\frac{1}{n}\right) \right] = \left(\frac{1}{n} \sum_i h(x_i)\right) - h\left(\frac{1}{n}\right)$$

and similarly

$$H_\phi(y,\ w) = \frac{1}{n} \sum_i h(w_i) - h\left(\frac{1}{n}\right).$$

Therefore

$$\frac{1}{n} \sum_i h(w_i) - h\left(\frac{1}{n}\right) \le \overline{\alpha}\left(\frac{1}{n} \sum_i h(x_i)\right) - \overline{\alpha} h\left(\frac{1}{n}\right)$$

or

$$\sum_i h(w_i) \le \overline{\alpha} \sum_i h(x_i) + \alpha n h\left(\frac{1}{n}\right).$$

Because h is convex,

$$\sum_i h(x_i) \ge n h\left(\frac{1}{n}\right)$$

hence

$$\overline{\alpha} \sum_i h(x_i) + \alpha\, n h\left(\frac{1}{n}\right) \le \sum_i h(x_i). \qquad \square$$

COROLLARY 2.2. *Under the hypotheses of Theorem 2.1, if in addition* $h(1/n) \le 0$, *then*

$$\sum_i h(w_i) \le \overline{\alpha} \sum_i h(x_i).$$

EXAMPLE 2.3. For $s \in [0, \infty)$, if $h(s) = s \log s$ (with $0 \log 0 = 0$, $x \in P_n$, $w = Ax$ and A any doubly stochastic $n \times n$ matrix,

$$\sum_i w_i \log w_i \le \overline{\alpha}(A) \sum_i x_i \log x_i - \alpha(A) \log n \le \overline{\alpha}(A) \sum_i x_i \log x_i .$$

EXAMPLE 2.4. If the hypotheses of Theorem 2.1 are weakened only by permitting h to be non-convex, then the conclusion no longer follows. Here is an example of a doubly stochastic A, x and $w = Ax$ such that $\sum h(x_i) \ge \sum h(w_i) > \overline{\alpha}(A) \sum h(x_i)$. For small $\varepsilon > 0$, let $x^T = (1 - \varepsilon, \varepsilon)$ and let

$$A = \begin{pmatrix} 1/4 & 3/4 \\ 3/4 & 1/4 \end{pmatrix}.$$

Then $w = Ax = (1/4 + \varepsilon/2, 3/4 - \varepsilon/2)^T$ and $\overline{\alpha}(A) = 1/2$. Define $h(t) = 0$ for $t \in (0, 1/8)$, $h(t) = 3/8$ for $t \in [1/8, 7/8]$, $h(t) = 1$ for $t \in (7/8, 1)$. Then $\sum h(x_i) = 1$, $\sum h(w_i) = 3/4$, as claimed.

3. Quantitative majorization in continuous processes

Analogous results hold for continuous-time processes. For background, see Alberti and Uhlmann (1982, pp. 30-31). Assume now that all matrices are $n \times n$ and real. A matrix in which all off-diagonal elements are nonnegative and the sum of every column is zero is called an intensity matrix; such matrices have zero or negative elements on the main diagonal. If B is an intensity matrix, it is well known that for all nonnegative real t, e^{Bt} is column-stochastic. A matrix in which all off-diagonal elements are nonnegative and the sum of every column and row is zero is called a double intensity matrix; in this case, for all nonnegative real t, e^{Bt} is doubly stochastic.

For $x(0) \in P_d$, $y(0) \in P_d$, and $t \ge 0$, define $x(t) = e^{Bt}x(0)$ and $y(t) = e^{Bt}y(0)$. The following is an immediate result of combining Theorems 4.1 and 7.1 and Corollary 7.3 of Cohen, Iwasa, Rautu, Ruskai, Seneta and Zbaganu (in press).

THEOREM 3.1. *If* B *is an intensity matrix,* $\omega \ge \max_i |b_{ii}|$, *and*

$$\alpha \equiv \alpha(\omega^{-1}B + I),$$

then

$$\frac{d}{dt} \log\{H_\phi(x(t), y(t))\} \le -\omega\alpha \le -\beta,$$

where

$$\beta \equiv \min_{j,k:\, j<k} \left\{ \sum_{\substack{i \ne j \\ i \ne k}} [\min(b_{ij}, b_{ik})] + b_{jk} + b_{kj} \right\}.$$

Because $\omega^{-1}B + I$ is column-stochastic, α is meaningful; moreover, α is positive, as previously noted, if $\omega^{-1}B + I$ is scrambling.

THEOREM 3.2. *Let $h : \Re \to \Re$ be a convex function, B a double intensity $n \times n$ matrix, $x(0) \in P_n$, $x(t) = e^{Bt}x(0)$, $t \geq 0$. Then*

$$\frac{d}{dt}\log\left[\sum_i h(x_i(t)) - nh\left(\frac{1}{n}\right)\right] \leq -\omega\alpha \leq -\beta,$$

where ω, α and β are defined in Theorem 3.1.

Proof. Choose $y(0)$ to be the n-vector with each element $1/n$. Then $y(t) = e^{Bt}y(0) = y(0)$. Define ϕ as in the proof of Theorem 2.1, so that, as in that proof,

$$H_\phi(y(t),\, x(t)) = \frac{1}{n}\left[\sum_i h(x_i(t)) - nh\left(\frac{1}{n}\right)\right].$$

Taking the logarithm, then the derivative, of both sides of this equation and applying Theorem 3.1 gives the desired result. □

An obvious corollary is

$$\frac{d}{dt}\sum_i h(x_i(t)) \leq 0,$$

which is well-known. Alberti and Uhlmann (1982, p. 30) state the result and cite earlier sources.

4. A converse for the monotonicity of relative entropy

The nonquantitative monotonicity theorem of relative entropy states that if A is a column-stochastic, row-allowable $m \times n$ matrix and x and y are positive n-vectors, then $H_\phi(Ax, Ay) \leq H_\phi(x, y)$ (e.g., Moran 1961, Csiszár 1963 [p. 90, his Theorem 1], Morimoto 1963). In this theorem, x and y need not be normalized to be probability n-vectors, whereas in the quantitative monotonicity theorem (Theorem 1.4), it is assumed that x and y are probability vectors. We now establish a converse of the nonquantitative monotonicity theorem of relative entropy with the additional hypothesis that x and y are probability vectors. This converse is the analogue, for the monotonicity of relative entropy, of the Hardy-Littlewood-Pólya-Karamata converse of Schur's theorem for majorization.

THEOREM 4.1. *Let $u, v \in P_m$, $x, y \in P_n$ be fixed. If $H_\phi(u, v) \leq H_\phi(x, y)$ for every relative ϕ-entropy H_ϕ that satisfies Definition 1.2, then there exists a row-allowable column-stochastic matrix A such that $u = Ax$ and $v = Ay$.*

The proof of Theorem 4.1 depends on Theorem 4.2, a fundamental result of Choquet theory due to Cartier, Fell and Meyer (1964). For other statements of Choquet theory and Theorem 4.2, see e.g. Winkler (1985) and Bratteli and Robinson (1987, ¶4.2.1). Let X be a compact metric space. Let S be a convex cone of measurable functions $f: X \to \Re$ such that the closure $\overline{S}$ of S (in the uniform topology) is closed under the max operation, i.e., if $f, g \in \overline{S}$, then $\max(f, g) \in \overline{S}$. Let $\mathrm{B}(X)$ be the family of Borel sets of X. A transition measure $T: X \times \mathrm{B}(X) \to \Re$, written as $T(x, dy) = T_x(dy)$, is defined to be a dilation if

$$f(x) \leq T_x(f) \equiv \int_{y \in X} f(y) T_x(dy) \quad \forall f \in S.$$

THEOREM 4.2. *Under the conditions on X and S just stated, if* μ *and* ν *are two positive measures on* $(X, \mathrm{B}(X))$ *such that*

$$\mu(f) \le \nu(f) \quad \forall f \in S,$$

where e.g. $\mu(f) \equiv \int_X f(x)\mu(dx)$*, then there exists a dilation T such that* $\nu = \mu T$*, i.e., for every bounded measurable f,* $\nu(f) = \mu(Tf)$*, i.e.,*

$$\nu(f) \equiv \int_X f(x)\nu(dx) = \mu(Tf) \equiv \int_X T_x(f)\mu(dx) = \int_X \int_X f(y)T_x(dy)\mu(dx).$$

We also require a simple lemma. Define

$$S = \{f : f:(0,\infty)\times(0,\infty) \to \Re, f \text{ is convex and homogeneous}\}$$

and $S_0 = \{f \in S : f(1,1) = 0\}$. S contains all linear functions $f(s,t) = as + bt$.

LEMMA 4.3. *Let* $u, v \in P_m$, $x, y \in P_n$. *If* $H_\phi(u,v) \le H_\phi(x,y)$ *for every* $\phi \in S_0$*, then* $H_f(u,v) \le H_f(x,y)$ *for every* $f \in S$.

Proof. For any $f \in S$, define $\phi(x,y) \equiv f(x,y) - f(x,x) = f(x,y) - xf(1,1)$. Then $\phi \in S_0$ and $H_\phi(u,v) = H_f(u,v) - f(1,1)$, $H_\phi(x,y) = H_f(x,y) - f(1,1)$. Hence $H_f(u,v) \le H_f(x,y)$. □

PROOF OF THEOREM 4.1. In our case, $X = [0,1]\times[0,1]$. Let S and S_0 be the restrictions to X of the convex cones defined just before Lemma 4.3. Both S and S_0 are closed under the max operation; in fact, they are even closed. Given fixed $u, v \in P_m$, $x, y \in P_n$, we shall suppose that all the points (x_j, y_j) are distinct. (If they are not, a slight modification of the proof is required, which the reader can supply.) Then define

$$\mu = \sum_{i=1}^{m} \varepsilon_{u_i, v_i}, \qquad \nu = \sum_{j=1}^{n} \varepsilon_{x_j, y_j},$$

where ε_{x_1, x_2} denotes the Dirac needle function, i.e., the point measure concentrated on $(x_1, x_2) \in X$. (Thus by definition $\varepsilon_{x_1,x_2}(f) = f(x_1, x_2)$.) Then

$$\mu(f) = \sum_{i=1}^{m} f(u_i, v_i), \quad \nu(f) = \sum_{j=1}^{n} f(x_j, y_j) .$$

The hypothesis of Theorem 4.1, $H_\phi(u, v) \le H_\phi(x, y)$ for every relative ϕ-entropy H_ϕ of Definition 1.2, may be expressed as $\mu(\phi) \le \nu(\phi)$ for all $\phi \in S_0$. By Lemma 4.3, it follows that $\mu(f) \le \nu(f)$ for all $f \in S$. By Theorem 4.2, there exists a dilation T such that $\nu(f) = \mu(Tf)$, i.e., for every bounded measurable $f : X \to \Re$,

$$\sum_{j=1}^{n} f(x_j, y_j) = \sum_{i=1}^{m} T_{u_i, v_i}(f) \tag{1}$$

Define the set $E \equiv \{(x_j, y_j) : 1 \le j \le n\}$. When f is $1_{X\text{-}E}$, the indicator of $X - E$, we get from (1)

$$\sum_{i=1}^{m} T_{u_i, v_i}(X - E) = 0.$$

Thus the support of the measures T_{u_i, v_i} is included in the finite set E. Define

$$a_{ij} = T_{u_i, v_i}(\{(x_j, y_j)\}), \ 1 \le i \le m, \ 1 \le j \le n .$$

Then

(2)
$$T_{u_i, v_i} = \sum_{j=1}^{n} a_{ij} \, \varepsilon_{x_j, y_j} .$$

When f is the indicator of the set $\{(x_j, y_j)\}$, (1) becomes

$$1 = \sum_{i=1}^{m} T_{u_i, v_i}(\{(x_j, y_j)\}) = \sum_{i=1}^{m} a_{ij} .$$

Since this holds for every $j = 1, ..., n$, the $m \times n$ matrix A with elements a_{ij} is column-stochastic.

Since T is a dilation, we have by definition

(3)
$$f(u_i, v_i) \le T_{u_i, v_i}(f) \ \ \forall f \in S .$$

From (2)

(4)
$$f(u_i, v_i) \le \sum_{j=1}^{n} a_{ij} f(x_j, y_j) \ \ \forall f \in S .$$

In particular, if $f(s, t) = as + bt$, then $f \in S$ and $-f \in S$ and (4) becomes the equality

(5)
$$au_i + bv_i = \sum_{j=1}^{n} a_{ij}(ax_i + by_i) .$$

For $a = 1, b = 0$, (5) becomes $u = Ax$. For $a = 0, b = 1$, (5) becomes $v = Ay$. In short, if $H_\phi(u, v) \le H_\phi(x, y)$ for all $\phi \in S_0$, then there must exist a column-stochastic matrix A such that $u = Ax$ and $v = Ay$.

Because u and v are positive, A is necessarily row-allowable. □

ACKNOWLEDGMENTS. We thank an anonymous referee for careful and helpful comments, and Mr. and Mrs. William T. Golden for hospitality to J.E.C. This collaboration originated in part as a result of the exchange program between the U.S. National Academy of Sciences and the Romanian Academy (1985) and in part as a result of the conference in Blaubeuren, Germany (1991) to honor Wolfgang Doeblin.

REFERENCES

Alberti, P. M. and Uhlmann, A. 1982, *Stochasticity and Partial Order: Doubly Stochastic Maps and Unitary Mixing*. Berlin: VEB Deutscher Verlag der Wissenschaften; Dordrecht and Boston: D. Reidel.

Bratteli, O. and Robinson, D. W. 1987 *Operator Algebras and Quantum Statistical Mechanics 1: C*- and W*-Algebras, Symmetry Groups, Decomposition of States*, 2d ed. New York, Berlin: Springer-Verlag.

Cartier, P., Fell, J. M. G., and Meyer, P.-A. 1964. Comparaison des mesures portées par un ensemble convexe compact. Bull. Soc. Math. France 92:435-445.

Cohen, J. E., Iwasa, Y., Rautu, Gh., Ruskai, M. B., Seneta, E., and Zbaganu, Gh. In press, Relative entropy under mappings by stochastic matrices. *Linear Algebra and Its Applications.*

Csiszár, I., 1963, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 8:85-108. Math. Rev. 29:333, #1671, 1965.

Csiszár, I. and Körner, J. 1981, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press; Budapest: Akademiai Kiado.

Dobrushin, R. L., 1956, Central limit theorem for nonstationary Markov chains. I, *Theory of Probability and Its Applications* 1:65-80; II, *ibid.* 1:329-383.

Doeblin, W., 1937, Le cas discontinu des probabilités en chaîne. *Publ. Fac. Sci. Univ. Masaryk (Brno)*, no. 236.

Hajnal, J., 1958, Weak ergodicity in non-homogeneous Markov chains, *Proc. Camb. Phil. Soc.* 54:233-246.

Hardy, G. H., Littlewood, J. E., and Pólya, G. 1929, Some simple inequalities satisfied by convex functions. *Messenger of Mathematics* 58:145-152.

Hardy, G. H., Littlewood, J. E., and Pólya, G., *Inequalities*, 2nd ed., Cambridge University Press, 1952.

Karamata, J. 1932, Sur une inégalité relative aux fonctions convexes. *Publ. math. Univ. Belgrade* 1:145-148.

Liese, F. and Vajda, I. 1987, *Convex Statistical Distances*, Leipzig: Teubner.

Marshall, A. W. and Olkin, I. 1979, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York.

Moran, P. A. P., 1961, Entropy, Markov processes and Boltzmann's H-theorem, *Proc. Camb. Phil. Soc.* 57:833-842.

Morimoto, T., 1963, Markov processes and the H-theorem. *Journal of the Physical Society of Japan* 18:328-331.

Roberts, A. W. and Varberg, D. E. 1973, *Convex Functions*, New York: Academic Press.

Schur, I. 1923, Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinanten-Theorie. *Sitzungsber. d. Berlin. Math. Gesellschaft* 22:9-20. *Issai Schur Gesammelte Abhandlungen* A. Brauer and H. Rohrbach, eds. 2:416-427. Springer-Verlag, New York, 1973.

Seneta, E. 1973, On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Camb. Philos. Soc.* 74:507-513.

Winkler, Gerhard. 1985 *Choquet Order and Simplices, with Applications in Probabilistic Models.* Lecture notes in mathematics, 1145. Berlin: Springer-Verlag.

ROCKEFELLER UNIVERSITY, 1230 YORK AVENUE, BOX 20, NEW YORK, NY 10021-6399, U.S.A.

E-mail address: cohen@rockvax.rockefeller.edu

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCES, 6 AV. V. LE GORGEU, 29287 BREST, FRANCE

CENTRE OF MATHEMATICAL STATISTICS, BD. MAGHERU 22, RO-70158, BUCURESTI, ROMANIA