*The Annals of Probability*
1987, Vol. 15, No. 3, 1217–1225

# MAJORIZATION, RANDOMNESS AND DEPENDENCE FOR MULTIVARIATE DISTRIBUTIONS[1]

BY HARRY JOE

*University of British Columbia*

The preorder relation of Hardy, Littlewood and Pólya (1929), Day (1973) and Chong (1974, 1976) is applied to multivariate probability densities. This preorder, which is called majorization here, can be interpreted as an ordering of randomness. When used to compare multivariate densities with the same marginal densities, it can be interpreted as an ordering of dependence or conditional dependence. Results in Hickey (1983, 1984) and Joe (1985) are generalized. A relative entropy function is proposed as a measure of dependence or conditional dependence for multivariate densities with the same marginals.

**1. Introduction.** Day (1973) and Chong (1974, 1976) extend the Hardy–Littlewood–Pólya (1929) preorder relation to measurable functions on a finite measure space and obtain rearrangement theorems and inequalities. This preorder relation is a generalization of vector majorization [see Marshall and Olkin (1979)]. In this paper, we will call it majorization and apply it to multivariate probability densities on measure spaces which can have finite or infinite measure. The majorization ordering on densities has an interpretation as an ordering of randomness and extends ideas in Hickey (1983, 1984). When used to compare multivariate densities with the same marginal densities, the majorization ordering can be interpreted as an ordering of dependence or conditional dependence. This generalizes the matrix majorization in Joe (1985), which can be interpreted as an ordering of dependence for contingency tables or for discrete bivariate distributions. We will show that a relative entropy function can be used as a measure of dependence or conditional dependence for multivariate densities with the same marginals.

In Section 2, we state several equivalent definitions of majorization for nonnegative integrable functions on a measure space, and generalize some results in Hickey (1983, 1984). In Section 3, the measure is taken to be a product measure and results are obtained for densities which are minimal with respect to the majorization ordering over a class of densities with the same marginals.

**2. Majorization for densities on a measure space.** Let $(X, \Lambda, \mu)$ be a measure space. For most applications, $X$ will be $\mathscr{R}^p$ or a countable subset of $\mathscr{R}^p$, and $\mu$ will be Lebesgue measure or counting measure. For a nonnegative integrable function $h$ on $(X, \Lambda, \mu)$, let $m_h(t) = \mu(\{x\colon h(x) > t\})$, $t \geq 0$, and $h^*(u) = m_h^{-1}(u) = \sup\{t\colon\ m_h(t) > u\}, 0 \leq u \leq \mu(X)$; $h^*$ is the (left-continuous)

Received June 1985; revised July 1986.

[1]Research supported by the Natural Sciences and Engineering Research Council of Canada through grant A-8698.

*AMS* 1980 *subject classifications*. Primary 62H20, 62H99.

*Key words and phrases*. Majorization, ordering of dependence, entropy.

decreasing rearrangement of $h$. The following theorem will be used to define majorization.

**Theorem 2.1.** *Let $f$ and $g$ be nonnegative integrable functions on $(X, \Lambda, \mu)$ such that $\int f\,d\mu = \int g\,d\mu$. The following are equivalent.*

(a) $\int [f - t]^+ \,d\mu \le \int [g - t]^+ \,d\mu$ *for all* $t \ge 0$, *where* $[y]^+ = \max(y, 0)$.

(b) $\int \phi(f)\,d\mu \le \int \phi(g)\,d\mu$ *for all convex, continuous real-valued functions* $\phi$ *with domain including the ranges of $f$ and $g$ such that* $\phi(0) = 0$ *and the integrals exist.*

(c) $\int_t^\infty m_f(s)\,ds \le \int_t^\infty m_g(s)\,ds$ *for all* $t \ge 0$.

(d) $\int_0^t f^*(u)\,du \le \int_0^t g^*(u)\,du$ *for all* $0 \le t < \mu(X)$.

**Proof.** (b) implies (a) because $\phi(u) = (u - t)^+$ is continuous and convex and satisfies $\phi(0) = 0$. We show that (a) implies (b) next. If $\phi'(0+) > -\infty$, then there is an increasing sequence of convex polygons $\phi_n(u) = \phi'(0+)u + \sum_{i=1}^n a_{in}(u - t_{in})^+$ with $a_{in} > 0$, $t_{in} > 0$, such that $\phi_n$ converges to $\phi$ on an interval including the ranges of $f$ and $g$. Condition (a) implies that $\int \phi_n(f)\,d\mu \le \int \phi_n(g)\,d\mu$. By the monotone convergence theorem, $\int \phi(f)\,d\mu \le \int \phi(g)\,d\mu$. If $\phi'(0+) = -\infty$, then let

$$\phi_n(u) = \begin{cases} \phi(u), & u \ge n^{-1}, \\ \phi(n^{-1}) + \phi'(n^{-1}-)(u - n^{-1}), & u < n^{-1}, \end{cases}$$

$n = 1, 2, \ldots$. Let $\psi_n(u) = \phi_n(u) - \phi_n(0)$. $\psi_n$ is convex, $\psi_n(0) = 0$ and $\psi_n'(0+) = \phi'(n^{-1}-)$. $\int \psi_n(f)\,d\mu$ exists because

$$\int \psi_n(f)\,d\mu = \int_{\{f \ge n^{-1}\}} \psi_n(f)\,d\mu + \int_{\{f < n^{-1}\}} \phi'(n^{-1}-) f\,d\mu.$$

By the previous argument, $0 \le \int [\psi_n(g) - \psi_n(f)]\,d\mu$. By the Lebesgue dominated convergence theorem, $\int [\psi_n(g) - \psi_n(f)]\,d\mu \to \int [\phi(g) - \phi(f)]\,d\mu$, using $\tau(g) + \tau(f)$ as the dominating function for all $n \ge j$, where $\tau(u) = |\phi(u)| + |\phi_j(0)| I_{[c,\infty)}(u)$, $j$ is a large integer and $c$ is equal one-half the positive root of $\phi$ if it exists or infinity if $\phi < 0$. Hence $\int \phi(f)\,d\mu \le \int \phi(g)\,d\mu$.

(a) is equivalent to (c) because $\int_X \psi(f)\,d\mu = -\int_0^\infty \psi(s)\,dm_f(s)$ for all real-valued Borel measurable functions $\psi$ (such that the integrals exist); letting $\psi(u) = (u - t)^+$ and using integration by parts, $\int [f - t]^+\,d\mu = -\int_t^\infty (s - t)\,dm_f(s) = \int_t^\infty m_f(s)\,ds$. See Section 1 of Chong (1974) for details.

The equivalence of (c) and (d) is Theorem 1.6 of Chong (1974). □

From now on we assume that all functions on $(X, \Lambda, \mu)$ are bounded and integrate to 1, so that they can be regarded as probability densities with respect to $\mu$. Also all sets that are used will be measurable.

**Definition 2.2.** Let $f$ and $g$ be densities on $(X, \Lambda, \mu)$. We say that $f$ is majorized by $g$ (denoted by $f \prec g$) if (a), (b), (c) or (d) of Theorem 2.1 holds.

**REMARK.** Vector majorization is usually defined as $x = (x_1, \ldots, x_n) \prec (y_1, \ldots, y_n) = y$ if $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ and $\sum_{i=1}^k x_i^* \le \sum_{i=1}^k y_i^*$, $1 \le k \le n$, where $x^*$ and $y^*$ are the decreasing rearrangement of $x$ and $y$. This is condition (d) with $X = \{1, 2, \ldots, n\}$, $f = x$, $g = y$ and $\mu$ = counting measure. Other special cases of Definition 2.2 are majorization for infinite sequences [Markus (1964); Marshall and Olkin (1979), page 16], continuous majorization for integrable functions on $[0, 1]$ [Ryff (1963, 1965)] and $p$-majorization [Cheng (1977); Marshall and Olkin (1979), Section 14A].

As in Hickey (1983, 1984), the majorization ordering can be interpreted as an ordering of randomness or uncertainty with $f$ being "more random" than $g$ if $f \prec g$. Let $A$ be a subset of $X$ with finite measure. Then, in the class of densities with support contained in $A$, the "most random" density is the uniform density on $A$.

**THEOREM 2.3.** *If $f(x) = (1/\mu(A))I_A(x)$, where $I_A$ is the indicator function of the set $A$, and $g(x)$ is a density satisfying $\int_A g\,d\mu = 1$, then $f \prec g$.*

**PROOF.** It is easy to verify that condition (a) of Theorem 2.1 holds. □

As in vector majorization, $f$ will be majorized by $g$ if $g$ is "averaged" in some way to get $f$. Examples of averaging operators are:

1. Let $A \in \Lambda$ and
$$f(x) = \begin{cases} g(x), & x \notin A, \\ \dfrac{1}{\mu(A)} \displaystyle\int_A g\,d\mu, & x \in A. \end{cases}$$
Then $f \prec g$ can be verified using condition (a).
2. Suppose $g(x) \le a < b \le g(y)$ for all $x \in A$, $y \in B$, where $A$ and $B$ are disjoint sets with $\mu(A) = \mu(B)$. Let $\varepsilon = (b - a)/2$. Then $f \prec g$ if
$$f(x) = \begin{cases} g(x), & x \notin A \cup B, \\ g(x) + \varepsilon, & x \in A, \\ g(x) - \varepsilon, & x \in B. \end{cases}$$
This can be verified using condition (a).
3. Suppose $k$: $X \times X \to [0, \infty)$ is doubly stochastic, i.e., $\int k(x, y)\,d\mu(y) = 1$ for all $x \in X$ and $\int k(x, y)\,d\mu(x) = 1$ for all $y \in X$. If $f(x) = \int k(x, y)g(y)\,d\mu(y)$, then $f \prec g$. This can be verified using condition (b) and Jensen's inequality. Suppose $X = \mathscr{R}^p$ and $g$, $h$ and $f$ are the densities of the random variables $Z_1$, $Z_2$ and $Z_1 + Z_2$, respectively, where $Z_1$ and $Z_2$ are independent. Letting $k(x, y) = h(x - y)$, we see that the density of $Z_1 + Z_2$ is majorized by that of $Z_1$ [cf. Hickey (1983, 1984)].

Following Hickey (1983, 1984), we introduce the concept of an uncertainty parameter.

DEFINITION 2.4. Let $\{f_\theta\}$ be a family of densities and $a(\theta)$ be a real-valued function. $a(\theta)$ is an uncertainty parameter if $f_\theta \prec f_{\theta'}$ whenever $a(\theta) \le a(\theta')$ [or $f_\theta \prec f_{\theta'}$ whenever $a(\theta) \ge a(\theta')$].

EXAMPLES. (i) Consider the family of uniform densities, i.e., for $A$ such that $\mu(A) < \infty$, $f_A(x) = (1/\mu(A))I_A(x)$. $f_{A_1} \prec f_{A_2}$ if $\mu(A_1) \ge \mu(A_2)$, so that $\mu(A)$ is an uncertainty parameter.

(ii) Let $X = \mathscr{R}^p$, $p \ge 2$, and $\mu$ = Lebesgue measure. Let $\psi$: $[0,\infty) \to [0,\infty)$ be a strictly decreasing function such that $\int_0^\infty \psi(y) y^{p/2-1}\,dy < \infty$. Consider the class $\{f_\Sigma\}$ of elliptically contoured densities [see, for example, Muirhead (1982)] with $f_\Sigma(x) = |\Sigma|^{-1/2}\psi(x\Sigma^{-1}x')$, where $x = (x_1,\dots,x_p)$, $x'$ is the transpose of $x$ and $\Sigma$ is a positive definite matrix. Then $f_{\Sigma_1} \prec f_{\Sigma_2}$ if $|\Sigma_1| \ge |\Sigma_2|$ so that $|\Sigma|$ is an uncertainty parameter. The special case of multivariate normal densities is obtained with $\psi(y) = (2\pi)^{-p/2}e^{-y/2}$.

PROOF OF (ii). Let $m_\Sigma(t) = \mu(\{f_\Sigma > t\})$. The formula for volume of the ellipsoid $\{x\colon x\Sigma^{-1}x' \le y\}$ is $\pi^{p/2}|\Sigma|^{1/2}y^{p/2}/\Gamma(p/2+1)$ so that

$$m_\Sigma(t) = \begin{cases} \pi^{p/2}|\Sigma|^{1/2}\left[\psi^{-1}\left(|\Sigma|^{1/2}t\right)\right]^{p/2}/\Gamma\left(\frac{p}{2}+1\right), & t \le |\Sigma|^{-1/2}\psi(0), \\ 0, & t > |\Sigma|^{-1/2}\psi(0). \end{cases}$$

$\int_t^\infty m_\Sigma(s)\,ds = \pi^{p/2}[\Gamma(p/2+1)]^{-1}\int_{t|\Sigma|^{1/2}}^{\psi(0)}[\psi^{-1}(s)]^{p/2}\,ds$ is decreasing as $|\Sigma|$ increases. The conclusion follows from condition (c) of Theorem 2.1. □

(iii) Consider the family of multinomial distributions with $p_1,\dots,p_k$ fixed $(\sum_{i=1}^k p_i = 1)$, i.e.,

$$f_n(x_1,\dots,x_k) = \binom{n}{x_1,\dots,x_k}\prod_{i=1}^k p_i^{x_i}.$$

$f_{n_1} \prec f_{n_2}$ if $n_1 > n_2$, so that $n$ is an uncertainty parameter. This follows from the result included with the third averaging operator.

## 3. Majorization for densities when marginals are fixed.

In this section, let $X = X_1 \times \cdots \times X_p$ be a product space and let $\mu = \mu_1 \times \cdots \times \mu_p$ be a product measure. For most applications $X_i$ will be $\mathscr{R}^{m(i)}$ or a countable subset of $\mathscr{R}^{m(i)}$ for some positive integer $m(i)$ and $\mu_i$ will be Lebesgue measure or counting measure. Let $\mu_{-i} = \times_{j\ne i}\mu_j$, $i = 1,\dots,p$. For $i = 1,\dots,p$, let $f_i(x_i)$ be a density on $X_i$ with respect to $\mu_i$. Let $\Pi = \Pi(f_1,\dots,f_p)$ be the class of densities $f$ on $(X, \Lambda, \mu)$ such that $\int f\,d\mu_{-i} = f_i$, $i = 1,\dots,p$. Consider the ordering in Section 2 to be constrained to $\Pi$.

DEFINITION 3.1. $f \prec g$ if there are marginal densities $f_1,\dots,f_p$ such that $f$, $g \in \Pi(f_1,\dots,f_p)$ and (a), (b), (c), or (d) of Theorem 2.1 holds.

A special case is the matrix majorization of Joe (1985), where $p = 2$, $X_1 = \{1, 2, \ldots, r\}$, $X_2 = \{1, 2, \ldots, c\}$, $r$ and $c$ are positive integers and $\mu_1$ and $\mu_2$ are counting measures. The matrix majorization ordering can be interpreted as an ordering of dependence for two-way contingency tables or for discrete bivariate distributions. This more general definition can be interpreted as an ordering of dependence for multivariate distributions with $g$ representing more "dependence" than $f$ if $f \prec g$. That is, with the marginals fixed, a multivariate distribution is more "random" if it is closer to "independence."

Now, some results in Joe (1985) will be generalized. A density $f$ in $\Pi$ is minimal if $f \succ g$ implies $f^* = g^*$. An approximate necessary condition (from Theorem 3.2 following) for $f$ to be minimal is that $f(x_1, \ldots, x_i, \ldots, x_p) \geq f(x_1, \ldots, x_i', \ldots, x_p)$ whenever $f_i(x_i) \geq f_i(x_i')$, $i = 1, \ldots, p$. Note that $f(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_i(x_i)$, which is the density of $p$ independent random variables, satisfies this condition. In general, there are also other densities that satisfy this condition.

**Theorem 3.2.** *Let $g \in \Pi$. If there are sets $A_1, \ldots, A_p$, $B_1, \ldots, B_p$ such that*

(i) $\mu_i(A_i) = \mu_i(B_i) > 0$, *$A_i$ and $B_i$ are disjoint*, $i = 1, \ldots, p$,

(ii) $f_1(x_1) \leq f_1(x_1')$ *for all* $x_1 \in A_1$, $x_1' \in B_1$,

(iii) $g(x_1, x_2) \leq a < b \leq g(x_1', x_2)$ *for all* $x_1 \in A_1$, $x_1' \in B_1$, $x_2 \in A_2 \times \cdots \times A_p = C_2$,

(iv) $g(x_1, x_2) \geq c > d \geq g(x_1', x_2)$ *for all* $x_1 \in A_1$, $x_1' \in B_1$, $x_2 \in B_2 \times \cdots \times B_p = D_2$,

*then $g$ is not minimal.*

**Proof.** Let $\varepsilon$ be the minimum of $(b - a)/2$ and $(c - d)/2$. Define $f(x)$ by

$$f(x) = \begin{cases} g(x), & x \notin (A_1 \cup B_1) \times (C_2 \cup D_2), \\ g(x) + \varepsilon, & x \in A_1 \times C_2 \cup B_1 \times D_2, \\ g(x) - \varepsilon, & x \in A_1 \times D_2 \cup B_1 \times C_2. \end{cases}$$

Then $f \in \Pi$ and $f \prec g$ (according to second averaging operators in Section 2). Hence $g$ is not minimal. □

A sufficient condition for $f \in \Pi$ to be minimal is given next.

**Theorem 3.3.** *Let $\phi$ be a continuous and strictly convex function with $\phi(0) = 0$. If $f \in \Pi$ minimizes $\int \phi(g)\, d\mu$, $g \in \Pi$, then $f$ is minimal.*

**Proof.** Suppose $f$ is not minimal, then there exist $g \in \Pi$ such that $g \prec f$ and $g^* \neq f^*$. By condition (b) of Theorem 2.1, $\varepsilon g + (1 - \varepsilon) f \prec f$ for all $0 \leq \varepsilon < 1$. Let $J(\varepsilon) = \int \phi(f + \varepsilon(g - f))\, d\mu$, $0 \leq \varepsilon \leq 1$. Then $J(\varepsilon) \leq J(0)$. Also $J(\varepsilon)$ is strictly convex so that $J(\varepsilon_0) < J(0)$ for some $0 \leq \varepsilon_0 < 1$. This contradicts the assumption that $f$ minimizes $\int \phi(g)\, d\mu$. □

REMARK. Minimizing $\int\phi(g)\,d\mu$ subject to the constraint $g \in \Pi$ is a calculus of variations type of problem. An argument similar to the preceding proof can be used to show that if a minimum exists, then it is unique (up to sets of measure zero). A necessary and sufficient condition for $f$ to be the minimizing function is given next.

THEOREM 3.4. *Suppose $\phi$ is differentiable and strictly convex with $\phi(0) = 0$ and the minimum of $\int\phi(g)\,d\mu$ exists. Then $f$ minimizes $\int\phi(g)\,d\mu$, $g \in \Pi$, if and only if $\int\phi'(f)h\,d\mu = 0$ for all integrable $h$ satisfying $\int h\,d\mu_{-i} = 0$, $i = 1,\dots,p$.*

PROOF. Suppose $f \in \Pi$ is the minimum. Note that if $h$ satisfies the previous condition, then $f + \varepsilon h \in \Pi$ (provided $f + \varepsilon h \geq 0$). Let $J(\varepsilon, h) = \int\phi(f + \varepsilon h)\,d\mu$ for $\varepsilon$ in a neighborhood of 0. Then $(\partial/\partial\varepsilon)J(\varepsilon, h) = \int\phi'(f + \varepsilon h)h\,d\mu$. Since $\varepsilon = 0$ minimizes $J(\varepsilon, h)$, $\int\phi'(f)h\,d\mu = 0$.

If $f$ satisfies the condition then $f$ is a local maximum or minimum. By convexity, $f$ is global minimum. □

COROLLARY 3.5. *If $\int\prod f_i \log\prod f_i$ exists, then $f(x_1,\dots,x_p) = \prod_{i=1}^p f_i(x_i)$ is minimal.*

PROOF. Let $\phi(u) = u\log u$, $u \geq 0$ $(0\log 0 = 0$ by convention). $\phi$ is strictly convex and $\phi'(u) = 1 + \log u$. It is easy to verify that the condition of Theorem 3.4 holds. □

COROLLARY 3.6. *If $\mu_i(X_i) = m_i < \infty$ and*

$$f(x) = m^{-1}\left[\sum_{i=1}^{p} m_i f_i(x_i) - (p-1)\right] \geq 0,$$

*where $m = \prod_{i=1}^p m_i$, then $f$ is minimal.*

PROOF. Let $\phi(u) = u^2$, $u \geq 0$. $\phi$ is strictly convex. It is easy to verify that the condition of Theorem 3.4 holds and that $f \in \Pi$. □

Corollary 3.5 suggests the use of the relative entropy function

$$\delta(g) = \int g\log\left[g/\prod f_i\right]d\mu = \int g\log g\,d\mu - \int\prod f_i(x_i)\log\prod f_i(x_i)\,d\mu$$

as a measure of dependence for $g \in \Pi(f_1,\dots,f_p)$. (The right hand equality depends on the existence of the integrals.) $\delta(g)$ is increasing with respect to the majorization ordering and takes a minimum of 0 when $g$ is the density representing independence. $\delta(g)$ can be used for $p > 2$; measures of dependence such as the correlation coefficient, sup correlation [Gebelein (1941); Rényi (1959)], monotone correlation [Kimeldorf and Sampson (1978)] are for bivariate distributions and do not generalize to multivariate distributions. We also note that $\int[g - \prod f_i]^2\,d\mu$, the measure of deviation from independence in Gilula and

Schwarz (1985), and the generalization of mean square contingency of Rényi (1959) are not increasing with respect to the majorization ordering. In the bivariate case, $\delta^*(g) = [1 - e^{-2\delta(g)}]^{1/2}$ [cf. Linfoot (1957)] satisfies conditions B, C, D, F, G of Rényi (1959) [see also Schweizer and Wolff (1981)] and it satisfies condition E for continuous random variables. $\delta(g)$ has the following nice property, which is a multivariate generalization of condition F in Rényi (1959): If the (continuous) random variables $V_1, \dots, V_p$ have joint density $g$ and marginal densities $f_1, \dots, f_p$, respectively, and $W_i = \tau_i(V_i)$ for differentiable one-to-one functions $\tau_i$, then $\delta(g) = \delta(h)$, where $h$ is the joint density of $W_1, \dots, W_p$. □

Next we introduce the concept of a dependence parameter.

DEFINITION 3.7. Let $\{f_\theta\}$ be a family of densities in $\Pi$ and let $a(\theta)$ be a real-valued function. $a(\theta)$ is a dependence parameter if $f_\theta \prec f_{\theta'}$ whenever $a(\theta) \le a(\theta')$ [or $f_\theta \prec f_{\theta'}$ whenever $a(\theta) \ge a(\theta')$].

EXAMPLES. (i) Let $\{f_\Sigma\}$ be the family of multivariate normal densities with $f_i(x_i) = (2\pi)^{-1/2}\sigma_i^{-1}\exp\{-x_i^2/2\sigma_i^2\}$, $i = 1, \dots, p$; that is, the diagonal entries of $\Sigma$ are $\sigma_1^2, \dots, \sigma_p^2$. $f_{\Sigma_1} \prec f_{\Sigma_2}$, if $|\Sigma_1| \ge |\Sigma_2|$ [from Example (ii) in Section 2], so that $|\Sigma|$ is a dependence parameter. Note that $|\Sigma|$ is maximized when $\Sigma = \operatorname{diag}(\sigma_1^2, \dots, \sigma_p^2)$, and $\delta^*(f_\Sigma) = (1 - |\Sigma|/\prod\sigma_i^2)^{1/2}$.

(ii) Let $R$ be a correlation matrix. The multivariate Student $t$ density with $n$ degrees of freedom is $f_R(x) = |R|^{-1/2}\psi_n(xR^{-1}x')$, where

$$\psi_n(y) = \Gamma((p+n)/2)(n\pi)^{-p/2}[\Gamma(n/2)]^{-1}(1 + y/n)^{-[(p+n)/2]}.$$

The marginal densities are univariate Student $t$ densities with $n$ degrees of freedom. By Example (ii) of Section 2, $|R|$ is a dependence parameter.

(iii) Let $\{f_\Sigma\}$ be the family of $m$-variate normal densities. Let $m = m_1 + m_2$, where $m_2 \ge m_1 \ge 1$, and let $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where $\Sigma_{jj}$ are fixed $m_j \times m_j$ positive definite matrices, $j = 1, 2$. Let $p = 2$ and let the marginal density $f_j$ be $m_j$-variate normal with covariance matrix $\Sigma_{jj}$, $j = 1, 2$. Similar to Example (i), $|\Sigma|$ is a dependence parameter and is maximized when $\Sigma_{12} = 0$. Furthermore

$$\delta^*(f_\Sigma) = \left(1 - \frac{|\Sigma|}{|\Sigma_{11}|\,|\Sigma_{22}|}\right)^{1/2} = \left[1 - \prod_{i=1}^{m_1}\left(1 - \rho_i^2\right)\right]^{1/2},$$

where $\rho_1 \ge \cdots \ge \rho_{m_1} \ge 0$ are the canonical correlation coefficients [see, for example, Muirhead (1982), page 531]. When $m_1 = 1$ with $\Sigma_{11} = \sigma_{11}$, $\delta^*(f_\Sigma) = (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}/\sigma_{11})^{1/2}$ is the square root of the multiple correlation coefficient from regressing the variable with density $f_1$ on the $m_2$ variables with joint density $f_2$.

The preceding results can be generalized to the case where higher-dimensional marginals are fixed. For example, consider the class $\Pi(f_{12}, f_{13})$ of trivariate densities $g$ with fixed bivariate marginals $f_{12} = \int g\,d\mu_3$, $f_{13} = \int g\,d\mu_2$. The condition in Theorem 3.4 becomes $\int \phi'(f)h\,d\mu = 0$ for all integrable $h$ satisfying

$\int h \, d\mu_2 = 0$, $\int h \, d\mu_3 = 0$. If $\phi(u) = u \log u$ and the minimum of $\int \phi(g) \, d\mu$, $g \in \Pi(f_{12}, f_{13})$, exists, then it occurs for

$$f(x_1, x_2, x_3) = \frac{f_{12}(x_1, x_2) f_{13}(x_1, x_3)}{f_1(x_1)} I(f_1(x_1) > 0),$$

where $f_1 = \int f_{12} \, d\mu_2 = \int f_{13} \, d\mu_3$. This density represents conditional independence of the second and third variables given the first. $\delta(g) = \int g \log[g f_1 / f_{12} f_{13}] \, d\mu$ is a measure of conditional dependence. If $f$ is a $m$-variate ($m \geq 3$) normal density and $f_{12}$ and $f_{13}$ are the marginal densities when the $m$th and $(m-1)$st variables are, respectively, integrated out, then $\delta^*(f) = [1 - e^{-2\delta(f)}]^{1/2}$ is the absolute value of the partial correlation coefficient of the $(m-1)$st and $m$th variables conditional on the first $m-2$ variables.

We end this section by briefly mentioning densities which are "large" with respect the majorization ordering. When $\mu_i$ are Lebesgue measures, maximal densities in $\Pi$ do not exist because the "most dependent" distributions with marginals $f_1, \ldots, f_p$ have singular components (Theorem 3.8 following). In the case $p = 2$, $X_i$ are finite sets and $\mu_i$ are counting measures, Joe (1985) obtains results for maximal densities; the number of maximal densities is finite. When $p \geq 3$ and $X_i$ are finite sets, the number of maximal densities can be infinite (unpublished results of the author).

**THEOREM 3.8.** *Let $f \in \Pi$. Let $\mu$ be Lebesgue measure. If there are sets $A_1, \ldots, A_p$, $B_1, \ldots, B_p$ such that*

(i) *$\mu_i(A_i) = \mu_i(B_i) > 0$, $A_i$, $B_i$ disjoint, $i = 1, \ldots, p$,*

(ii) *$0 < \varepsilon_1 \leq f(x_1, x_2) \leq \varepsilon_2$ for all $x_1 \in A_1$, $x_2 \in C_2 = A_2 \times \cdots \times A_p$ and $x_1 \in B_1$, $x_2 \in D_2 = B_2 \times \cdots \times B_p$,*

(iii) *$f(x_1, x_2) \geq \varepsilon_2$ for all $x_1 \in A_1$, $x_2 \in D_2$ and $x_1 \in B_1$, $x_2 \in C_2$,*

*then $f$ is not maximal (there exists $g \in \Pi$ such that $f \prec g$ and $f^* \neq g^*$).*

**PROOF.** Let

$$g(x) = \begin{cases} f(x), & x \notin (A_1 \cup B_1) \times (C_2 \cup D_2), \\ f(x) - \varepsilon_1, & x \in A_1 \times C_2 \cup B_1 \times D_2, \\ f(x) + \varepsilon_1, & x \in A_1 \times D_2 \cup B_1 \times C_2. \end{cases}$$

Then $g \in \Pi$ and $f \prec g$. □

## REFERENCES


CHENG, K. W. (1977). Majorization: Its extensions and preservation theorems. Technical Report No. 121, Dept. Statistics, Stanford Univ.

CHONG, K.-M. (1974). Some extensions of a theorem of Hardy, Littlewood and Pólya and their applications. *Canad. J. Math.* **26** 1321–1340.

CHONG, K.-M. (1976). Doubly stochastic operators and rearrangement theorems. *J. Math. Anal. Appl.* **56** 309–316.

DAY, P. W. (1973). Decreasing rearrangements and doubly stochastic operators. *Trans. Amer. Math. Soc.* **178** 383–392.

GEBELEIN, H. (1941). Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.* **21** 364–379.

GILULA, Z. and SCHWARZ, G. (1985). On the maximum of a measure of deviation from independence between discrete random variables. *Ann. Probab.* **13** 314–317.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1929). Some simple inequalities satisfied by convex functions. *Messenger Math.* **58** 145–152.

HICKEY, R. J. (1983). Majorisation, randomness and some discrete distributions. *J. Appl. Probab.* **20** 897–902.

HICKEY, R. J. (1984). Continuous majorisation and randomness. *J. Appl. Probab.* **21** 924–929.

JOE, H. (1985). An ordering of dependence for contingency tables. *Linear Algebra Appl.* **70** 89–103.

KIMELDORF, G. and SAMPSON, A. R. (1978). Monotone dependence. *Ann. Statist.* **6** 895–903.

LINFOOT, E. H. (1957). An informational measure of correlation. *Inform. and Control* **1** 85–89.

MARKUS, A. S. (1964). Eigenvalues and singular values of the sum and product of linear operators. *Uspekhi Mat. Nauk* **19** 93–123.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10** 441–451.

RYFF, J. V. (1963). On the representation of doubly stochastic operators. *Pacific J. Math.* **13** 1379–1386.

RYFF, J. V. (1965). Orbits of $L^1$ functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* **117** 92–100.

SCHWEIZER, B. AND WOLFF, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.* **9** 879–885.

DEPARTMENT OF STATISTICS
2021 WEST MALL
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA V6T 1W5
CANADA