Stochastic Inequalities
IMS Lecture Notes – Monograph Series
Volume 22 (1993)

GENERALIZED MAJORIZATION ORDERINGS AND APPLICATIONS

By HARRY JOE[1]

University of British Columbia

Orderings that are special cases of or related to the majorization ordering in Joe (1987a) for functions on a measure space are reviewed. Applications in probability and statistics that have motivated the orderings are briefly discussed and some new applications are given. Also some new links are made between results of previous papers.

1. Introduction

This article reviews a class of majorization orderings that generalize vector majorization and some applications motivating or coming from the orderings. The emphasis is on work that has come after the publication of Marshall and Olkin (1979). The class fits within the majorization ordering in Joe (1987a) for functions on a measure space and includes most generalized majorization orderings. Exceptions are group majorization (see Eaton (1987), Giovagnoli and Wynn (1985)) and stochastic majorization (see Shanthikumar (1987)).

In Section 2, the definition of Joe (1987a) is given and then it is shown how other orderings are either special cases or are related in some way. A diversity of applications are discussed or summarized in Section 3. Marshall and Olkin (1979) unified inequalities through majorization, and although generalized majorization leads to inequalities, they have not always been the motivation for extensions. It is hoped that the results in this paper will lead readers to discover further applications and extensions.

2. Generalized Majorization Orderings

The goal in this section is to show that results of various authors fit within a unified framework. These authors have often not cross-referenced each other. We start with the definition of Joe (1987a).

[1]Research supported by NSERC Canada Grant A–8698.

AMS 1991 *subject classifications.* Primary 06A06, 62A99.

Key words and phrases. Entropy, majorization, ordering, probability distribution.

Let $(\mathcal{X}, \Lambda, \nu)$ be a measure space. For most applications, $\mathcal{X}$ will be a subset of a Euclidean space, and ν will be Lebesgue measure or counting measure. For a nonnegative integrable function h on $(\mathcal{X}, \Lambda, \nu)$, let $m_h(t) = \nu(\{x : h(x) > t\})$, $t \geq 0$, and let $h^*(u) = m_h^{-1}(u) = \sup\{t : m_h(t) > u\}$, $0 \leq u \leq \nu(\mathcal{X})$; h^* is the (left-continuous) decreasing rearrangement of h.

DEFINITION 2.1 Let a and b be nonnegative integrable functions on $(\mathcal{X}, \Lambda, \nu)$ such that $\int a\, d\nu = \int b\, d\nu$. Then a is majorized by b, written $a \prec b$, if one of the following four equivalent conditions hold.

(a) $\int [a-t]^+ d\nu \leq \int [b-t]^+ d\nu$ for all $t \geq 0$, where $[y]^+ = \max\{y, 0\}$.

(b) $\int \psi(a) d\nu \leq \int \psi(b) d\nu$ for all convex, continuous real-valued functions ψ such that $\psi(0) = 0$ and the integrals exist.

(c) $\int_t^\infty m_a(s)\, ds \leq \int_t^\infty m_b(s)\, ds$ for all $t \geq 0$.

(d) $\int_0^t a^*(u)\, du \leq \int_0^t b^*(u)\, du$ for all $0 \leq t < \nu(\mathcal{X})$.

ASIDE 2.2 The ideas in Definition 2.1 go back to Hardy, Littlewood and Pólya (1929) for $\mathcal{X}$ being a finite interval of the real line and to Chong (1974) for the general case. Except in some cases such as Case 1 and Case 2 (with $-\infty < C < D < \infty$) when $\mathcal{X}$ is a bounded subset of the real line, it is not the same as the dilation ordering of measures given in Chapter 13 of Phelps (1966) and references therein.

The above definition is suitable for all stochastic applications except one in this article; in the exception, a, b can be partly negative. If one wants to compare a, b that can have negative parts, then it appears necessary to have $\nu(\mathcal{X})$ finite. Note that $\int [a-t]^+ d\nu$ is not defined for an integrable function a if $t < 0$ and $\nu(\mathcal{X}) = \infty$.

If the $\nu(\mathcal{X})$ is finite and the nonnegativity condition for a, b is removed, then (i) m_a, m_b can be defined on $\mathbb{R}$, (ii) in condition (a), the inequality holds for all $t \in \mathbb{R}$, and (iii) $\psi(0) = 0$ in (b) is not required. In this case, the equivalence of parts (a), (c) and (d) still follows from Chong (1974). The main new contribution in Joe (1987a) is stating together the equivalent conditions as a generalized majorization ordering, and including condition (b) which is crucial to the applications in that paper. For $\nu(\mathcal{X})$ finite, another equivalent condition is

(a') $\int (a-t)^- d\nu \leq \int (b-t)^- d\nu$ for all real t, where $(y)^- = \max\{0, -y\}$.
The proof of condition (b) from (a') and (a) then follows with a few steps. The general case will follow after proving condition (b) for ψ such that $\psi'(0+) = 0 = \psi(0)$. The case where ψ has domain on $(-\infty, 0]$ can be handled similarly to when the domain is $[0, \infty)$ (see the proof of Theorem 2.1 in Joe (1987a)). If ψ has domain $[C, D]$ with $C < 0 < D$, then $\psi(x)$ can be approximated from below by functions of the form $\sum_{s_m < \cdots < s_1 < 0} (s_i - x)^+ + \sum_{0 < t_1 < \cdots < t_n} (x - t_i)^+$ and then the monotone convergence theorem can be used.

The following are specific cases (of Definition 2.1) that have been studied.

Case 1. The usual vector majorization results if $\mathcal{X} = \{1, \ldots, n\}$ and $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$ and ν is counting measure. Condition (d) in Definition 2.1 is the usual definition with decreasing components of the vectors a, b. The generalization to Case 2 below comes from thinking of a, b as probability vectors; i.e., a_i, b_i are probability masses at a point x_i. The generalization to Case 3 below comes from considering random variables X and Y with masses n^{-1} at points a_i and b_i respectively.

Case 2. Continuous majorization for densities on an interval $[C, D]$ (possibly unbounded) results if $a = f$, $b = g$, where f, g are densities on $[C, D]$ with respect to Lebesgue measure. In this case, $\prec$ is an ordering of closeness to uniformity of the density, and it has been studied in Hickey (1984). This ordering on densities can be extended to higher dimensional Euclidean space. In related work, Chan, Proschan and Sethuraman (1987), following up on Ryff (1963), study the majorization ordering for integrable functions (not necessarily nonnegative) on [0,1].

Case 3. The Lorenz ordering in Arnold (1987) and Das Gupta and Bhansali (1989) results if $\mathcal{X} = [0, 1]$, ν is Lebesgue measure and $a = F^{-1}$, $b = G^{-1}$, where F, G are the respective distribution functions of nonnegative random variables X, Y (X and Y have a common finite mean) and F^{-1}, G^{-1} are the corresponding quantile functions. Since a, b are monotone increasing, the decreasing rearrangements are respectively $a^*(u) = F^{-1}(1-u)$ and $b^*(u) = G^{-1}(1-u)$. Condition (d) of Definition 2.1 is $\int_0^t F^{-1}(1-p)\,dp \le \int_0^t G^{-1}(1-p)\,dp$, $0 < t < 1$ or

$$\text{(2.1)} \qquad \int_0^t F^{-1}(p)\,dp \ge \int_0^t G^{-1}(p)\,dp, \qquad 0 < t < 1;$$

the latter being the definition in Arnold (1987) and Das Gupta and Bhansali (1989). Condition (b) becomes

$$\text{(2.2)} \qquad \mathrm{E}\,\psi(X) \le \mathrm{E}\,\psi(Y) \qquad \forall \text{ convex continuous functions } \psi$$

and condition (a) becomes

$$\text{(2.3)} \qquad \mathrm{E}\,(X-t)^+ \le \mathrm{E}\,(Y-t)^+, \qquad \forall t \ge 0.$$

The Lorenz ordering is known by other names in earlier work. Let $\overline{F} = 1 - F, \overline{G} - 1 - G$ be the survival functions of F, G. In the form of condition (c) of Definition 2.1, that is,

$$\text{(2.4)} \qquad \int_u^\infty \overline{F}(x)dx \le \int_u^\infty \overline{G}(x)dx \qquad \forall \text{ real } u,$$

the ordering is referred to as the "more variable" ordering in Ross (1983) and as a majorization ordering in Boland and Proschan (1986).

If the random variables X, Y have support on a bounded interval, say [0,1], then the doubly stochastic condition of vector majorization generalizes in two ways. From Ryff (1965), Definition 2.1 for $F^{-1} \prec G^{-1}$ in this case is equivalent to the condition:

(e) there exists a doubly stochastic operator T (a positive, contraction operator such that $\int_0^1 (TI_E)d\nu = \nu(E)$, where ν is Lebesgue measure and I_E is the indicator function of the measureable set E) from L^1 to L^1 such that $F^{-1} = TG^{-1}$, where L^1 is the space of Lebesgue integrable functions on [0,1].

Also, from Theorem 10 of Blackwell (1951), it is equivalent to the condition:

(f) there is a stochastic transformation $H(x|y)$ ($H(\cdot|y)$ is a distribution for each y in the support of G) such that $\int H(x|y)\,dG(y) = F(x)$ for all x and $E(Y|X) = X$.

Note that if X, Y have support on the points $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ respectively with masses n^{-1} at each support point, then both conditions (e) and (f) are equivalent to the existence of a doubly stochastic matrix P such that $(x_1, \ldots, x_n) = (y_1, \ldots, y_n)P$.

Case 4. Simonis (1988) defines a "spectral order" which is the Lorenz ordering in Case 3 without the constraint of nonnegativity on the random variables X, Y (see also Aside 2.2). Let X, Y have respective distribution functions F, G, and corresponding survival functions $\overline{F}, \overline{G}$. Simonis defines $X \prec Y$ if $\mathrm{E}X = \mathrm{E}Y$ and $\int_0^t \overline{F}^{-1}(u)du \le \int_0^t \overline{G}^{-1}(u)du$ for all $0 < t < 1$ This is the same as (2.1) and condition (d) of Definition 2.1 but without the nonnegativity requirement for $a = F^{-1}$ and $b = G^{-1}$. Simonis proves the (a), (b) and (c) are equivalent, that is, $X \prec Y$ if and only if (2.3), (2.2) or (2.4) hold.

This ordering is also used in Stoyan (1983), where it is called a convex ordering or an ordering of "mean residual life" (with (2.3) as the definition).

Case 5. Non-uniform weighted majorization results if a measure other than Lebesgue measure or counting measure is used. For the vector case, let $\mathcal{X} = \{1, \ldots, n\}$ and ν be a measure with positive mass q_i at the point i, $i = 1, \ldots, n$. Cheng (1977) defined this ordering for vectors that are similarly ordered and called it p–majorization. Joe (1990) uses this ordering without the constraint of similarly ordered and in addition used the continuous version, that is, with $\mathcal{X} = [C, D]$ being an interval of the real line and ν corresponding to a positive density $q(\cdot)$ on $[C, D]$. Both of these orderings will be referred to as majorization with respect to q and denoted by $\prec_q$. For this case, the various forms of Definition 2.1 and other equivalent conditions are given and discussed in Joe (1990).

Case 6. The r–majorization ordering with respect to q of Joe (1990), denoted by $\prec_q^r$, follows from Case 5, with ratios of densities with respect to q. That is, for n–dimensional probability vectors $p_1 = (p_{11}, \ldots, p_{1n})$ and

$p_2 = (p_{21}, \ldots, p_{2n})$, $p_1 \prec_q^r p_2$ if $(p_{11}/q_1, \ldots, p_{1n}/q_n) \prec_q (p_{21}/q_1, \ldots, p_{2n}/q_n)$, and for densities p_1, p_2 of random variables on $[C, D]$, $p_1 \prec_q^r p_2$ if $p_1/q \prec_q p_2/q$. R-majorization puts the probability vector or density q at the lower end of the ordering instead the the uniform vector or density, and can be interpreted as an ordering of divergence or distance from q. With applications to thermodynamics, this ordering is called a mixing distance in Ruch and Mead (1976) and Ruch, Schranner and Seligman (1978).

3. Applications

In this section, we summarize some recent applications of the majorization orderings in Section 2, and also we give some new applications (in 3.5 and 3.9 and part of 3.2). In some cases, the application motivated the study of the majorization ordering. One goal is to show a diversity of applications so they are mainly brief (in which case details can be found in the papers that are referred to). The applications taking up more space are the new ones and the one in 3.8 which is more detailed in order to mention an open problem.

3.1. *Ordering on Random Variables or Cumulative Distribution Functions*

The ordering in Case 3 of Section 2 has been applied in diverse areas. Applications in reliability are given in Boland and Proschan (1986) and Ross (1983) with earlier such applications going back to Marshall and Proschan (1970). Applications to queueing models are given in Ross (1983) and Stoyan (1983). Arnold (1987) has applications to distributions of wealth; the idea is that with a given mean wealth μ, a distribution of wealth F with a constant F^{-1} is most equitable (this corresponds to a mass of 1 at μ) and a distribution F which is larger in the (Lorenz) ordering is less equitable. A further, more detailed application is given next.

3.2. *Probability Forecasting*

Conditions (e) and (f) of Section 2 are used in this application. DeGroot and Fienberg (1982) and DeGroot and Erikkson (1985) consider an ordering of forecasters with possible forecasts in the set $\{x_0, x_1, \ldots, x_m\}$, where $x_0 = 0, x_m = 1$ and $x_0 < x_1 < \cdots < x_m$. Their ordering can be generalized to allow forecasts in [0,1], in which case, their ordering becomes that in Application 3.1 or Case 3 of Section 2 with support of random variables in the interval [0,1].

The framework is that forecasters give a probability each day for an event like occurrence of rain. A forecast of 1 means a prediction that the event will happen and a forecast of 0 means a prediction that the event will not happen. A forecaster is well-calibrated if the conditional probability of the event, given that the forecaster's prediction is x, is x. Let two well-calibrated forecasters A, B have distributions F, G for their forecasts. Then the means of F, G must be the same, both being the probability or relative frequency of the event. Forecaster B is at least as refined as A if condition (f) holds. In well-behaved situations, the doubly stochastic operator in conditions (e) and (f) becomes an integral operator, that is, there is a function k on $[0,1]^2$ such that $\int k(u,v)du = 1$ for all v, $\int k(u,v)dv = 1$ for all u, and $\int k(u,v)G^{-1}(v)\,dv = F^{-1}(u)$. In fact, $k(u,v) = h(F^{-1}(u)|G^{-1}(v))/f(F^{-1}(u))$, where $h(\cdot|y)$ is the density of $H(\cdot|y)$ and f is the density of F; both densities are with respect to a measure that dominates F (for simple cases, Lebesgue measure, counting measure or a combination).

Let p be the relative frequency of the event. It is intuitively true, and not difficult to show from Definition 2.1, that the best or most refined well-calibrated forecaster has a forecast distribution G that has mass of p at 1 and mass of $1-p$ at 0. The least refined well-calibrated forecaster has a forecast distribution F that has a mass of 1 at p. With the restriction that forecasts are in the set $\{x_0, x_1, \ldots, x_m\}$, referred to above, DeGroot and Fienberg (1982) proved that the least refined well-calibrated forecaster has a distribution F that puts mass α at x_i and mass $1-\alpha$ at x_{i+1}, where i is such that $x_i < p \le x_{i+1}$ and $\alpha = (x_{i+1} - p)/(x_{i+1} - x_i)$.

3.3. *Constrained Majorization*

The author has used constrained majorization in several papers: Joe (1985, 1987a,b, 1988a, 1990). This comes about when there are additional constraints on a, b (in Definition 2.1) of the form

$$\int h_\tau a\,d\nu = \int h_\tau b\,d\nu, \quad \forall \tau \in \Upsilon,$$

where Υ is an index set. Only functions that satisfy the additional constraints are comparable. With constrained majorization, maximal and minimal functions in the orderings can be of interest.

3.4. *Orderings of Dependence*

If the h_τ are taken to be appropriate indicator functions and $a = f$, $b = g$ are m-dimensional multivariate densities with respect to a measure ν, the constraints can be that f, g have the same set of univariate margins, say $f_1, \ldots, f_m$. Then the constrained majorization ordering is an ordering of dependence among densities in the set $\Gamma(f_1, \ldots, f_m)$ of multivariate densities with univariate margins $f_1, \ldots, f_m$. This and generalizations are studied in Joe (1987a) and Joe (1985), with the latter mainly concerned with the bivariate discrete case with counts from a two-way contingency table (see also Application 3.6) as well as bivariate distributions. With this ordering, the density $f_I = \prod_j f_j$ is among those which are minimal. If it is desired to have f_I as the "unique" minimal density, then ratios relative to f_I can be used with Case 6 of Section 2 (that is, r-majorization with respect to f_I). The ordering of dependence in Scarsini (1990) is r-majorization for the bivariate case. However Scarsini also allows for the pair of univariate margins not being identical for comparison of two densities.

3.5. *Exploratory Data Analysis for Two-Way Tables*

There is a benefit to having more than just f_I as the minimal density in Application 3.4, especially for two-way tables of counts or sample proportions. For this special case, Joe (1985) proves that a necessary condition for minimal tables is that each row is similarly ordered with the column sum margin and each column is similarly ordered with the row sum margin. Since minimal tables with respect to the constrained majorization ordering can be interpreted as those "closest" to independence, this result provides a quick way to check whether two categorical variables are approximately independent. This can be done in one's head, unlike computation of the expected counts under the assumption of independence.

An example illustrating this is given below; the data are from students at the University of British Columbia in a recent year who took a first year calculus course. The two-way table below is constructed from the two variables, with grade in calculus (A, B or $\leq C$) as the column variable and type of high school (Vancouver, rest of Greater Vancouver Regional District, rest of British Columbia, private) as the row variable. The last row and the last column are the marginal totals by grade and by type of high school.

	A	B	$\leq C$	Total
1	198	143	201	542
2	186	169	284	639
3	80	102	159	341
4	42	40	83	165
Total	506	456	727	1687

By checking for similarly ordered rows and columns, one can see that the two variables are close to being independent, with the main discrepancies being that more A's than expected under independence are from "Vancouver" (compare first and fourth columns) and less A's than expected under independence are from "rest of B.C." (compare third and fifth rows). A measure of dependence given in Joe (1987a) is $\delta^* = (1 - e^{-2\delta})^{1/2}$, where

$\delta = \sum_{i,j} p_{ij} \log[p_{ij}/(p_{i+}p_{+j})]$, p_{ij} is the proportion simultaneously in category i of the row variable and category j of the column variable, $p_{i+} = \sum_j p_{ij}$ and $p_{+j} = \sum_i p_{ij}$. The value of δ^* is 0.13 for the above table, and this suggests very little dependence as δ^* can take values between 0 and 1.

3.6. *Fisher's Exact Test*

The maximal and minimal tables from the constrained majorization ordering in Applications 3.3 and 3.4 were used in the network algorithm of Mehta and Patel (1983) for the computation of the P–value of Fisher's exact test for two–way contingency tables. This allowed the use of maxima and minima of certain functions instead of the use of the bounds in Mehta and Patel. The theorems from Joe (1985) were programmed into a Fortran routine, and the improvement in computational time reported in Joe (1988b) and Clarkson, Fan and Joe (1990). The routine has now been adapted into IMSL and Splus.

3.7. *Ordering of Transitivity*

Depending on what $\mathcal{X}$ is, constrained majorization can lead to interpretations other than an ordering of dependence. An example is for paired comparison matrices $a = \{p_{ij} : 1 \leq i, j \leq n, i \neq j\}$, where there are n items and p_{ij} is the probability that item i is preferred to item j. The ordering on a with the constraints $\sum_{j \neq i} p_{ij} = m_i$, $i = 1, \ldots, n$, is interpreted as an ordering of transitivity in Joe (1988b), in that matrices at the lower end of the ordering are such that there is a preference transitivity among the items.

3.8. *Conjugate Priors and Majorization with Moment Constraints*

The ordering Case 6 was partly motivated with the aim to justify the use of some common conjugate priors. This goal was only partly reached and the remaining step to be proved is posed as a problem here. This application also shows the connection between majorization and entropy, and illustrates what constrained majorization results are like.

Consider the problem of choosing a prior distribution for a random (continuous) quantity after having elicited the mean and/or variance. We compare the use of r–majorization with the maximum entropy principle and the principle of minimum cross entropy (Jaynes 1983, Shore and Johnson 1980). The following results from Joe (1990) are needed; they concern minimal densities relative to r–majorization with respect to q when there are first and/or second moment constraints.

THEOREM 3.8.1 *Let $q(x)$ be a positive continuous function on the interval $[C, D]$. Let $\mathcal{P} = \mathcal{P}(C, D; \mu)$ be the class of densities f (with respect to Lebesgue measure) on the interval $[C, D]$ satisfying $\int_C^D x f(x) dx = \int_C^D x\, r(x) q(x)\, dx = \mu$, where $r(x) = f(x)/q(x)$. Then $f \in \mathcal{P}$ is minimal with respect to q if and only if $r(x)$ is monotone.*

THEOREM 3.8.2 *Let $q(x)$ be a positive continuous function on the interval $[C, D]$. Let $\mathcal{P} = \mathcal{P}(C, D; \mu_1, \mu_2)$ be the class of densities f (with respect to Lebesgue measure) on the interval $[C, D]$ satisfying $\int_C^D x f(x) dx = \mu_1$, $\int_C^D x^2 f(x) dx = \mu_2$. Then $f \in \mathcal{P}$ is minimal with respect to q only if $r(x) = f(x)/q(x)$ is monotone, U-shaped or unimodal.*

By taking a limit, Theorems 3.8.1 and 3.8.2 are valid for open (and possibly infinite) intervals and q can approach ∞ at one or both of the endpoints.

One approach to eliciting a prior distribution might be to first use the invariance principle to obtain a "non-informative", possibly improper, prior q (Cox and Hinkley (1974, Chapter 10)) and then use knowledge (possibly subjective) of first and/or second order moments to choose a prior close to q satisfying the moment constraints. The minimum cross entropy principle (and the maximum entropy principle as a special case) lead to a very small class of priors – for example, only the exponential and normal, possibly truncated, result as prior distributions from the maximum entropy principle. Although several common conjugate priors are maximum entropy based on other constraints, a stronger justification is based on first and/or second moment constraints since these are more easily elicited than something like the expected value of the logarithm of the quantity.

It is known (see, for example, Berger (1980)) that if the minimum cross entropy distribution with respect to q exists, then it has the form

$$f(x) = C(\lambda_1, \ldots, \lambda_m) q(x) \exp[-\sum_j \lambda_j h_j(x)], x \in \mathcal{X},$$

when the constraints are $\int_{\mathcal{X}} h_j(x) f(x) d\nu = \mu_j$, $j = 1, \ldots, m$. When q is not a constant function and the constraints are moments, no commonly used conjugate prior probability distribution has this form. It is shown below that several conjugate prior distributions of an unknown scale parameter σ satisfy the necessary conditions for minimality with respect to the invariant (improper) prior which is proportional to $q(\sigma) = 1/\sigma$. That is, the constrained r-majorization results lead to a larger class of distributions that are "near" q, subject to the moment constraints. Nearness of a density to q here means that for any closed subinterval of $(0, \infty)$ the density is (relatively) near c/σ, where c is the normalizing constant for the interval.

Conjugate priors for scale parameters include (a) the inverse gamma density, $f(\sigma) \propto \sigma^{-\alpha-1} \exp\{-\beta/\sigma\}$, $\sigma > 0$, for a gamma distibution with known shape parameter, (b) the density, $f(\sigma) \propto \sigma^{-2\alpha-1} \exp\{-\lambda/\sigma^2\}$, $\sigma > 0$, for a normal distribution with known mean, and (c) the hyperbolic or Pareto density (Raiffa and Schlaifer (1961)), $f(\sigma) \propto \sigma^{-\alpha-1}$, $\sigma \geq M$ $(M > 0)$, for a uniform distribution on 0 to an unknown upper bound. It is easy to check that $\sigma f(\sigma)$ is unimodal over $(0, \infty)$ for (a) and (b) and $\sigma f(\sigma)$ is monotone over $[M, \infty)$ for (c) when $\alpha > 0$. Hence the necessary condition for minimality in Theorem 3.8.2 is satisfied for (a) and (b), and the necessary condition for minimality in Theorem 3.8.1 is satisfied for (c). Note that $C(\lambda_1, \lambda_2)\sigma^{-1} \exp\{\lambda_1\sigma + \lambda_2\sigma^2\}$ from (3.1) is not a proper density on $(0, \infty)$.

For an example not involving a scale parameter, consider the probability parameter θ $(0 < \theta < 1)$ of a binomial distribution. "Noninformative" priors that have been proposed are $q(\theta) = [\theta(1-\theta)]^{-0.5}$ (Jeffreys (1961)), $q(\theta) = [\theta(1-\theta)]^{-1}$ (Haldane (1948)) and $q(\theta) \equiv 1$. The conjugate prior for the parameter θ is the Beta density $f(\theta)$, which is proportional to $\theta^{\alpha-1}(1-\theta)^{\beta-1}$. Hence, for all three of these q, $f(\theta)/q(\theta)$ is monotone, U–shaped or unimodal, and the necessary condition for minimality when there are two moment constraints is satisfied. This is better than saying that the beta distribution is a maximum entropy distribution subject to knowing the expectation of $\log\theta$ and $\log(1-\theta)$. It would be nice to prove in addition that the beta density is minimal relative to r–majorization with respect to q for any of the above q (since satisfying the necessary condition for minimality need not imply minimality). This however is an open problem (a similar comment holds for cases (a) and (b) in the preceding paragraph). The techniques in Joe (1990) do not work but numerical comparisons with the maximum entropy density on (0,1) show that the beta density with the same first two moments does not majorize the maximum entropy density.

3.9. *Updating Subjective Probability*

This is another application of r–majorization.

For simplicity of presentation, we think of a random quantity which has a finite number of possible outcomes, labelled as $1, 2, \ldots, n$; however results do generalize. Suppose our initial prior distribution is $q = (q_1, \ldots, q_n)$, with q_i being the probability of outcome i. If we then get further information that cause us to revise our probabilities for some pairwise mutually exclusive events, how should we update our probability distribution? Diaconis and Zabell (1982) study this problem using Jeffrey's conditionization rule and a divergence or distance approach; the latter was called mechanical updating. We look at this updating problem using majorization as an ordering of divergence instead of using several different measures of divergence. This is, in a sense, a simpler way of obtaining and viewing the results in Section

5 of Diaconis and Zabell.

Let $E_1, \ldots, E_e$ be pairwise mutually exclusive events (subsets of $S = \{1, \ldots, n\}$). Without loss of generality, we suppose that these form a partition of S. Let $q^* = (q_1^*, \ldots, q_n^*)$ be the updated subjective probability distribution, where our further information causes us to decide

$$q^*(E_j) = \sum_{i \in E_j} q_i^* = \eta_j, \qquad j = 1, \ldots, e, \tag{3.1}$$

where $\sum_j \eta_j = 1$. The constrained r–majorization approach is to consider the r–majorization ordering with respect to q ($\prec_q^r$) for the class of q^* that satisfy (3.1).

The minimal q^* is such that q^*/q is as close to uniform as possible, and therefore, from Joe (1990), it is piecewise uniform and satisfies $q_i^*/q_i = c_j$, $i \in E_j$, $j = 1, \ldots, e$, for some constants c_j, and from (3.1), clearly, $c_j = \eta_j / \sum_{i \in E_j} q_i$. Hence Jeffrey's rule,

$$q^*(A|E_j) = \frac{\sum_{i \in A \cap E_j} q_i^*}{\eta_j} = \frac{c_j \sum_{i \in A \cap E_j} q_i}{\eta_j} = \frac{\sum_{i \in A \cap E_j} q_i}{\sum_{i \in E_j} q_i} = q(A|E_j), \tag{3.2}$$

for the conditional probabilities holds. Theorem 5.1 of Diaconis and Zabell (1982) makes the conclusion (3.2) for the q^* that minimizes the Hellinger or cross entropy divergence from q subject to (3.1). This follows as a corollary of the r–majorization result since both of these divergence measures are increasing in the ordering $\prec_q^r$ (Section 3.1 of Joe, 1990).

Diaconis and Zabell (1982) also study the case of (compatible) updated probabilities for two partitions $E_1, \ldots, E_e$ and $D_1, \ldots, D_d$. Let q^* be the updated probability distribution. Suppose $q^*(E_j) = \eta_j$ and $q^*(D_k) = \theta_k$. Let $A_{jk} = E_j \cap D_k$. Note that A_{jk}, $j = 1, \ldots, e$, $k = 1, \ldots, d$, form a partition of S. Let $q(A_{jk}) = \zeta_{jk} = \sum_{i \in A_{jk}} q_i$ and let $q^*(A_{jk}) = \zeta_{jk}^* = \sum_{i \in A_{jk}} q_i^*$. Then the revised probabilites are specified up to

$$\sum_k \zeta_{jk}^* = \eta_j, \qquad j = 1, \ldots, e, \qquad \sum_j \zeta_{jk}^* = \theta_k, \qquad k = 1, \ldots, d.$$

Consider the set Q^* of q^* which satisfy these constraints and put the r–majorization ordering with respect to q on this set. Now there is not a unique minimal q^* in Q^* for this ordering. This explains why Diaconis and Zabell obtained different q^* for different divergence measures (cross entropy and variation distance) However, in an argument similar to the above, it can be concluded that minimal distributions in Q^* are piecewise uniform and must satisfy $q_i^*/q_i = \zeta_{jk}^*/\zeta_{jk} = c_{jk}$ for $i \in A_{jk}$ for some constants c_{jk}. If the ζ_{jk}'s are all the same, then similar to Theorem 1 of Joe (1985), a minimal q^* must be such that the matrix (c_{jk}) has rows which are similarly ordered to $(\theta_1, \ldots, \theta_d)$ and columns which are similarly ordered to $(\eta_1, \ldots, \eta_e)$. In general there is no simple characterization of the minimal q^*. But a sufficient condition for q^* in Q^* to be minimal is that it minimizes

$$\sum_i q_i \psi(q_i^*/q_i) = \sum_j \sum_k \zeta_{jk} \psi(\zeta_{jk}^*/\zeta_{jk})$$

for a strictly convex real-valued function ψ. If $\psi(u) = u \log u$, the minimum cross entropy distribution results and its form is $\zeta_{jk}^* = \zeta_{jk}\alpha_j\beta_k$ for some positive constants α_j, β_k. Hence $\zeta_{jk}^* = \eta_j\theta_k$ is a minimum cross entropy distribution only if $\zeta_{jk} = \zeta_{j+}\zeta_{+k}$ where $\zeta_{j+} = \sum_k \zeta_{jk}$ and $\zeta_{+k} = \sum_j \zeta_{jk}$. The conclusion here is not always the suggestion of updated probabilities, $\zeta_{jk}^* = \eta_j\theta_k$, in Section 4.2 of Diaconis and Zabell (1982).

Acknowledgements. I am grateful to the referees for comments leading to an improved presentation.

REFERENCES

ARNOLD, B. C. (1987). *Majorization and the Lorenz Order.* Springer–Verlag, New York.

BERGER, J. O. (1980). *Statistical Decision Theory.* Springer–Verlag, New York.

BLACKWELL, D. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Math. Statist. Prob.*, J. Neyman, ed. University of California Press, Berkeley, CA. 93–102.

BOLAND, P. J. AND PROSCHAN, F. (1986). An integral inequality with applications to order statistics. In *Reliability and Quality Control*, A.P. Basu, ed. North Holland, Amsterdam, 107–116.

CHAN, W. T., PROSCHAN, F. AND SETHURAMAN, J. (1987). Schur–Ostrowski theorems for functionals on $L_1(0,1)$. *SIAM J. Math. Anal.* **18** 566–578.

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

CLARKSON, D. B., FAN, Y.-A. AND JOE, H. (1990). A remark on algorithm 643: FEXACT: An algorithm for performing Fisher's exact test in $r \times c$ contingency tables. Tech. Report, Department of Statistics, University of British Columbia.

COX, D. R. AND HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

DAS GUPTA, S. AND BHANDARI, S.K. (1989). Multivariate majorization. In *Contributions to Probability and Statistics, Essays in honour of Ingram Olkin.* L. J. Gleser, M. D. Perlman, S. J. Press and A. R. Sampson, eds. Springer–Verlag, New York, 63–74.

DEGROOT, M. H. AND ERIKKSON, E. A. (1985). Probability forecasting, stochastic dominance, and the Lorenz curve. In *Bayesian Statistics* **2** J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. Smith, eds. North-Holland, Amsterdam, 99–118.

DeGroot, M. H. and Fienberg, S. E. (1982). Assessing probability assessors: calibration and refinement. In *Statistical Decision Theory and Related Topics III*, **Vol. 1**. S. S. Gupta and J. O. Berger, eds. Academic Press, New York, 291–314.

Diaconis, P. and Zabell, S. A. (1982). Updating subjective probability. *J. Amer. Statist. Assoc.* **77** 822–831.

Eaton, M. L. (1987). *Lecture on Topics in Probability Inequalities.* Centrum voor Wiskunde en Informatica, Amsterdam.

Giovagnoli, A. and Wynn, H. P. (1985). G–majorization with applications to matrix orderings. *Linear Alg. Appl.* **67** 111–135.

Haldane, J. B. S. (1948). The precision of observed values of small frequencies. *Biometrika* **35** 297–303.

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. (1984). Continuous majorisation and randomness. *J. Appl. Prob.* **21** 924–929.

Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkrantz, ed. Reidel, Dordrecht.

Jeffreys, H. (1961). *Theory of Probability,* 3rd edition. Oxford University Press, New York.

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

Joe, H. (1987a). Majorization, randomness and dependence for multivariate distributions. *Ann. Probab.* **15** 1217–1225.

Joe, H. (1987b). An ordering of dependence for distributions of k–tuples, with applications to lotto games. *Canad. J. Statist.* **15** 227–238.

Joe, H. (1988a). Majorization, entropy and paired comparisons. *Ann. Statist.* **16** 915–925.

Joe, H. (1988b). Extreme probabilities for contingency tables under row and column independence with application to Fisher's exact test. *Commun. Statist.-Theor. Meth.* **17** 3677–3685.

Joe, H. (1990). Majorization and divergence. *J. Math. Anal. Appl.* **148** 287–305.

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

Marshall, A. W. and Proschan, F. (1970). Mean life of series and parallel systems. *J. Appl. Prob.* **7** 165–174.

Mehta, C. R. and Patel, N. R. (1983). A network algorithm for performing Fiesher's exact test in $r \times c$ contingency tables. *J. Amer. Statist. Assoc.* **78** 427–434.

Phelps, R. R. (1966). *Lectures on Choquet's Theorem.* Van Nostrand, Princeton, NJ.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Colonial Press, Clinton, MA.

Ross, S. M. (1983). *Stochastic Processes.* Wiley, New York.

Ruch, E. and Mead, C. A. (1976). The principle of increasing mixing character and some of its consequences. *Theor. Chim. Acta.* **41** 95–117.

Ruch, E., Schranner, R. and Seligman, T. H. (1978). The mixing distance. *J. Chem. Phys.* **69** 386–392.

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.

SCARSINI, M. (1990). An ordering of dependence. In *Topics in Statistical Dependence.* H. W. Block, A. R. Sampson, and T. H. Savits, eds. Institute of Mathematical Statistics, Hayward, CA. 403–414.

SHANTHIKUMAR, J. G. (1987). Stochastic majorization of random variables with proportional equilibrium rates. *Adv. Appl. Prob.* **19** 854–872.

SHORE, J. E. AND JOHNSON, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross–entropy. *IEEE Trans. Inform. Theory* **IT-26** 26–37.

SIMONIS, A. (1988). Ordem Espectral e Schur-Convexidade em Espacos de Probabilidade. M.Sc. Thesis, Instituto de Mathematica e Estatistica, Universidade de Sao Paulo.

STOYAN, D. (1983). *Comparison Methods for Queues and Other Stochastic Models.* Wiley, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, B.C. CANADA V6T 1Z2