Stochastic Orders and Decision under Risk
IMS Lecture Notes – Monograph Series (1991)

REPRESENTATION THEOREMS FOR MEASURES OF LOCATION AND FOR MEASURES OF DISPERSION

By Alessandra Giovagnoli and Giuliana Regoli

University of Perugia

In this paper we characterize classes of statistical functionals through some results which have been inspired by a classical theorem on mean values due to de Finetti, Nagumo, and Kolmogorov. All the functionals are order preserving w.r.t. particular stochastic orderings.

The first result is a quasi-linear representation for functionals assuming, together with monotonicity w.r.t. first degree stochastic dominance and associativity, a particular continuity condition which can be interpreted as a mild type of robustness. This result is used in a "dual" way to characterize other measures of location, like median, quantiles, trimmed means, and Winsorized means.

In the second part of the paper our aim is to characterize some measures of dispersion of a distribution around its expected value which are order preserving w.r.t. the so-called dilation ordering. Most statistical indices of variability can be obtained in this way. This and a "dual" theorem also account for several measures of inequality, which are order preserving with respect to the concentration ordering based on the Lorenz curve, like Gini's celebrated index.

1. Introduction. In 1930 Nagumo and Kolmogorov (independently) and later de Finetti proved a classical theorem which characterizes suitable means of n real numbers (in Nagumo's and Kolmogorov's versions) and of distributions with support in an interval (in de Finetti's) as quasi-linear functions, in fact 1-1 transformations of expected values. In a decision theory context the dFNK theorem states necessary and sufficient conditions in order to represent the certainty equivalent of a monetary lottery as a transformed expected utility, where $u(\cdot)$, the utility function, is increasing. This theorem has been extended in many ways (see for instance, Chew, 1983 and 1989, and Fishburn, 1986).

A crucial assumption of dFNK is associativity which in the parallel development of expected utility theory is replaced by a stronger requirement, the

---

AMS 1980 Subject Classification: Primary 62C05; Secondary 90A43, 62F35.

Key words and phrases: Dilation ordering, L-functionals, Lorenz ordering, quasi-linear means, stochastic dominance.

so-called independence axiom, or sure thing principle. Another assumption is monotonicity w.r.t. the classical stochastic ordering, – also known as first degree stochastic dominance. Stronger orderings, like second and third degree stochastic dominance, often play an important role in economic applications (Fishburn and Vickson, 1978): the representation result of the dFNK theorem still applies, assuming that one such order is preserved, and indeed properties of the utility function $u(\cdot)$ related to the type of risk aversion can be specified better, namely second degree stochastic dominance implies that $u(\cdot)$ is concave, as well as increasing, and third degree stochastic dominance additionally that $u'(\cdot)$ is convex (see Giovagnoli and Regoli, 1990).

This paper focuses not on utility theory, but on more statistical applications of the dFNK theorem and modified versions thereof. We deal with general functionals of a distribution, to be thought of either as the probability or the frequency distribution of a character in a population, or as an empirical or sample distribution. The first result is a modification of the dFNK theorem in which a quasi-linear representation for such functionals is obtained assuming, together with associativity and monotonicity, a particular continuity condition which, as pointed out in Giovagnoli and Regoli (1990), for statistical and economic indices can be interpreted as a mild type of robustness.

The result is used in a "dual" way to characterize other measures of location, like median, quantiles, trimmed means, and Winsorized means. Roughly speaking, whereas quasi-linear means are – up to a transformation – the expected values of a transformed random variable, the other location measures we look at are expected values of a random variable w.r.t. transformed distributions. Our result is similar in flavor to one obtained by Yaari (1987) for representation of preferences. Here again we are concerned with the classical stochastic ordering, which is the natural order relation to consider for all measures of location, as well as for utilities.

In the second part of the paper our aim is to characterize some measures of dispersion. We give a representation theorem for classes of functionals measuring the dispersion of a distribution around its expected value which are order preserving w.r.t. the so-called dilation ordering, namely we express them as transforms of the expectation of a convex function. Most statistical indices of variability can be obtained in this way. In decisions under risk, such indices can be interpreted as measures of risk with respect to a convex loss function. This and a "dual" theorem also account for several measures of inequality, which are order preserving with respect to the concentration ordering based on the Lorenz curve, like Gini's celebrated index.

We would like to stress that what characterizes most of our results is the monotonicity of the functional w.r.t. a particular stochastic ordering. The different orderings under consideration, apart from playing a technical role in the proofs – namely obtaining an integral representation for the functional for distributions whose support is not finite – are what singles out and somehow defines each time the concept which the statistical indices aim to measure.

2. A de Finetti-Nagumo-Kolmogorov Type Theorem Under an Assumption of Continuity with Respect to Weights. The dFNK theorem as given by Hardy, Littlewood and Polya (1934), characterizes "means" of distributions with support contained in the interval $[a,b]$ which are associative, consistent, and strictly monotonic with respect to the stochastic ordering (first degree stochastic dominance) as functionals of the form

$$m(F) = u^{-1}\left(\int_a^b u(x)dF\right);$$

$u(\cdot)$ is continuous and strictly monotonic and is unique up to affine transformations.

We recall that *associativity* means that $m(F_1) = m(F_2) \Rightarrow m((1-\lambda)F_1 + \lambda G) = m((1-\lambda)F_2 + \lambda G)$ for all distributions F_1, F_2, G, and $\forall\ 0 \leq \lambda \leq 1$; *consistency* means that $m(\delta_x) = x\ \forall\ x \in [a,b]$, where δ_x denotes the distribution concentrated in x (point mass). The *stochastic ordering* $\leq^{st}$ is defined as $F \leq^{st} G$ iff $F(x) \geq G(x)\ \forall\ x \in [a,b]$.

A possible interpretation of the theorem is to view m not as a "mean", but simply as a functional Φ from a set of distribution functions to the real line: economic and statistical indicators are functionals of this type. The dFNK theorem allows one to say that under suitable assumptions the functional Φ is quasi-linear, i.e. $\exists$ a real $\phi(\cdot)$, invertible and such that $\Phi(\Sigma_i\lambda_i F_i) = \phi^{-1}(\Sigma_i\lambda_i\phi\Phi(F_i))$ for all $\lambda_i \geq 0$, $\Sigma_i\lambda_i = 1$, and to represent it as the transform of a sum or an integral. Associativity indicates that a given "mean value" of a character in a given population is left unaltered if part of the population is replaced by another of equal size having the same mean value as the replaced part. This property is shared by a large number of indices Φ, other than "means", and it often seems a fairly natural condition to assume for a statistical functional. Monotonicity w.r.t. $\leq^{st}$ is a natural requirement for a "mean", or a measure of location, and in general for all indicators of size, preference, utility, etc.

On the other hand, if we are speaking more in general of a functional Φ on the set of distributions with support in $[a,b]$ which does not necessarily take values in $[a,b]$ itself, consistency does not always make sense, or may not be of interest, while the following condition may be more appropriate:

CONDITION 1. The function $\Phi(\delta_x) = \phi(x)$ is continuous in $[a,b]$.

With this slight change, and under the same monotonicity and associativity assumptions, the dFNK theorem continues to hold, since ϕ is also invertible and $m(F) = \phi^{-1}\Phi(F)$ is an associative, monotonic and consistent mean; we only need to replace u^{-1} by $\psi = \phi u^{-1}$, namely

$$\Phi(F) = \psi\left(\int_a^b u(x)dF\right).$$

Condition 1 means that a small change in the value of the character, when this value is assumed with probability = 1, leads to a small change of the value of the index: a related condition is that of "continuity with respect to weights" given in Giovagnoli and Regoli (1990) and recalled below, which appears to have an appealing interpretation in statistical terms, namely that a change, even a large change, of the character in a tiny percentage of the population brings about a minimal change of the index.

We shall denote by $D(X)$ the set of distributions with support in a set $X \subseteq R$ and by $D(X)^0$ the distributions in $D(X)$ with finite support; S will be a convex set $\subseteq D(X)$ and $S^0 = S \cap D(X)^0$; X will be omitted form the notation if given once and for all.

DEFINITION. A functional $\Phi : S \to R$ is said to be *continuous in* $F \in S$ *with respect to variations of* a *(finite number) of weights* if

$$\forall\, G = \Sigma_i \lambda_i \delta_{x_i} \text{ in } S^0, \ \ \forall\, \epsilon > 0,\ \exists\, \eta_\epsilon \text{ such that if } 0 < \eta < \eta_\epsilon \text{ then}$$
$$|\Phi((1-\eta)F + \eta\Sigma_i\lambda_i\delta_{x_i}) - \Phi(F)| < \epsilon;$$

Φ is said to be *continuous w.r.t. weights* if it is continuous in every $F \in S$. It is convenient to write

$$\lim_{\eta\to 0} \Phi((1-\eta)F + \eta G) = \Phi(F) \tag{2.1}$$

Some simple algebra shows that if S^0 consists of all the distributions with finite support in X, continuity w.r.t. a single weight – i.e. $\forall\, F \in S, \forall\, x \in X$, $\forall\, \epsilon > 0$, $\exists\, \eta_\epsilon > 0$ such that $\forall\, 0 < \eta < \eta_\epsilon \Rightarrow |\Phi((1-\eta)F + \eta\delta_x) - \Phi(F)| < \epsilon$ – implies continuity w.r.t. several weights.

The continuity-with-respect-to-weights condition is a type of robustness condition for statistical indices. This condition appears to be a reasonable requirement in many cases and compares with the robustness conditions for indices mentioned by Bickel and Lehmann (1975). On the other hand, this continuity property is implied by, but is much weaker than, the classical concept of robustness according to Huber (1981), i.e. continuity with respect to weak convergence of the F's. For instance the expected value and the variance,, which are not robust when the support is unbounded, are nevertheless continuous with respect to weights. The continuity-with-respect-to-weights is however a necessary condition for the existence of the *influence curve*, defined by Hampel as the Gâteaux derivative of Φ in the direction δ_x, i.e.

$$\lim_{\eta \to 0} \frac{\Phi((1-\eta)F + \eta\delta_x) - \Phi(F)}{\eta}$$

(see Huber, 1981). One property of this type of continuity, which extends (2.1), is given in the following Lemma.

LEMMA 1. *Let $\Phi : S \to R$ be continuous in F with respect to weights. Then $\forall\, G \in S^0$ and $\forall\, \lambda_o$* with $0 \leq \lambda_o < 1$

$$\lim_{\lambda \to \lambda_o} \Phi((1-\lambda)F + \lambda G) = \Phi((1-\lambda_o)F + \lambda_o G). \tag{2.2}$$

PROOF.

$$\begin{aligned} &|\Phi((1-\lambda-\epsilon_1)F + (\lambda+\epsilon_1)G) - \Phi((1-\lambda)F + \lambda G| \\ &= |\Phi([1-(1-\lambda)^{-1}\epsilon_1][(1-\lambda)F + \lambda G)] \\ &\quad + (1-\lambda)^{-1}\epsilon_1 G) - \Phi((1-\lambda)F + \lambda G) \ < \epsilon \end{aligned}$$

by taking $\eta = (1-\lambda)^{-1}\epsilon_1$ in (2.1). ∎

Lemma 1 also implies that $|\Phi(F) - \Phi(G)| < \epsilon$ for two step functions F and G with the same support that differ only by a small amount on every step.

LEMMA 2. *If $\Phi : S \to R$ is associative and has a maximum and a minimum, the function $\psi(\lambda) = \Phi((1-\lambda)m + \lambda M)$, $0 \leq \lambda \leq 1$, where $\Phi(m) = \min\{\Phi(s) : s \in S\}$ and $\Phi(M) = \max\{\Phi(s) : s \in S\}$, does not depend on the choice of m and M and in $(0,1)$ is either constant or $1-1$.*

Furthermore, if Φ is continuous with respect to weights, the function $\psi(\lambda) = \Phi((1-\lambda)m + \lambda M)$ is continuous in λ.

PROOF. For the first part see de Finetti (1931) or Daboni and Wedlin (1982). The second part follows from Lemma 1. ∎

Lemma 2 is used for the proof of the following technical lemma, which utilizes traditional arguments like those of DeGroot (1970). A full proof is given in Giovagnoli and Regoli (1989).

LEMMA 3. *Let $S = H(E)$ be the convex hull of E, where $E = U_n E_n$, and $\{E_n\}$ is an increasing sequence. Let $\Phi : S \to R$, and assume Φ has a maximum and a minimum in each E_n, with $\Phi(m_n) = \min\{\Phi(e), e \in E_n\}$, $\Phi(M_n) = \max\{\Phi(e), e \in E_n\}$. Assume further that*

(1) Φ is associative,

(2) $\forall\ e \in E_n$, $\exists\ \lambda \in [0,1]$ such that $\Phi(e) = \Phi((1-\lambda)m_n + \lambda M_n)$ i.e. $\psi_n(\lambda) = \Phi((1-\lambda)m_n + \lambda M_n)$ maps $[0,1]$ onto $\Phi(E_n)$,

then $\exists$ a real function ψ, invertible, such that given $s = \Sigma_i\lambda_i e_i \in S$, $\Phi(s) = \psi(\Sigma_i\lambda_i\psi^{-1}\Phi(e_i))$.

We can now prove the following.

THEOREM 1. *Let $D(X)^B$ = all distributions with bounded support in a given set $X \subseteq R$. Then a functional $\Phi : D(X)^0 \to R$ is*

(1) associative,

(2) continuous w.r.t. weights,

if and only if $\exists$ two real functions $u : X \to [0,1]$ and ψ, where ψ is continuous and invertible, such that for all distributions with finite support

$$F = \Sigma_i\lambda_i\delta_{x_i},\ \ x_i \in X \Rightarrow \Phi(F) = \psi(\Sigma_i\lambda_i u(x_i));$$

$u(\cdot)$ is unique up to an affine transformation.

Furthermore, (1) and (2) hold and $\Phi : D(X)^B \to R$ is

(3) monotonic (not necessarily strictly) with respect to the stochastic ordering

$$\text{iff } \Phi(F) = \psi\left(\int_X u(x)dF\right) \ \text{ with } u(\cdot) \text{ monotonic.}$$

PROOF. We start by proving the quasi-linearity under conditions (1) and (2). Two sequences $\{N_n\}$ and $\{M_n\}$ can be found in $D(X)^0$ in such a way that $\forall\ F \in D^0$, $\Phi(N_n) \leq \Phi(F) \leq \Phi(M_n)$ for some n and such that the sequence $E_n = \{F \in D^0 : \Phi(F) \in [\Phi(N_n), \Phi(M_n)]\}$ is increasing. By Lemma 2, the functions $\psi_n(\lambda) = \Phi((1-\lambda)N_n + \lambda M_n)$ are continuous, and thus map $[0,1]$ onto $\Phi(E_n)$. Hence Lemma 3 can be applied with $S = E = D(X)^0$. For all $x \in X$ define $u(x) = \psi^{-1}\Phi(\delta_x)$ whence $\Phi(\Sigma_i\lambda_i\delta_{x_i})\psi(\Sigma_i\lambda_i(u(x_i))$. Since Φ is continuous with respect to weights, $\psi(\lambda)$ is continuous. It is straightforward to check that $\psi(\Sigma_i\lambda_i u(x_i)) = \phi(\Sigma_i\lambda_i v(x_i)) \Rightarrow v(\cdot) = au(\cdot) + b$, $a, b \in R$.

Now assume condition (3) holds too. The function $\psi(\cdot)$ is strictly monotonic since it is continuous and 1-1. Since by (3) $\Phi(\delta_x)$ is also monotonic in x, so is $u(x) = \psi^{-1}\Phi(\delta_x)$. We need to prove that the integral representation obtains. This can be done by the same argument used in the original proof of the dFNK theorem – see for instance Daboni and Wedlin (1982) – which only requires (non-strict) monotonicity w.r.t. the stochastic ordering. ∎

REMARK. In the proof of this theorem, condition (2) is required only to prove the continuity of $\psi(\lambda)$.

By this theorem conditions (1) and (2) characterize all the functionals on D^B which are 1-1 transforms of an expected utility, with a utility function $u(\cdot)$ that may or may not be continuous. Note that the theorem can be applied also when the distributions have support in a discrete set, e.g. $X = \{0, 1, 2, \cdots\}$. On the other hand the geometric mean, min(supp F), and max(supp F) are examples of functionals which are monotonic, not strictly, w.r.t. $\leq^{st}$, but to which Theorem 1 cannot be applied, since they are not continuous with respect to weights.

COROLLARY. *If X is connected, assuming continuity of Φ w.r.t. weak convergence in place of (2), the same result holds and moreover $u(\cdot)$ is continuous.*

PROOF. Continuity w.r.t. weak convergence implies (2). Continuity of $u(\cdot)$ follows from $u(x) = \psi^{-1}\Phi(\delta_x)$ and $x \to x_0 \Leftrightarrow \delta_x \xrightarrow{w} \delta_{x_0}$. ∎

3. Representing Other Measures of Location. Theorem 1 can be applied to the inverses of distribution functions, namely:

$$F^{-1}(p) = \inf\{x : F(x) > p\} \qquad p \in [0, 1)$$

to obtain a representation theorem for other "means", i.e. measures of location, which in a way possess some "dual" properties. What we have in mind in particular are the median and all the other quantiles, the trimmed means, and the Winsorized means; as is well known, these are all functionals of $F(x)$ of the form

$$\int x dh(F(x)) \tag{3.1}$$

with $h(\cdot)$ an increasing real function on $[0, 1]$, right continuous but not continuous. In fact, quantiles, trimmed means, and Winsorized means are obtained, respectively, when

$$\text{(i)} \qquad h(t) = \begin{cases} 0 & \text{for } t \in [0, \alpha), \\ 1 & \text{for } t \in [\alpha, 1]; \end{cases}$$

$$\text{(ii)} \qquad h(t) = \begin{cases} 0 & t \in [0, \alpha) \\ (t - \alpha)/(\beta - \alpha) & t \in [\alpha, \beta) \\ 1 & t \in [\beta, 1] \qquad 0 < \alpha < \beta < 1 \end{cases} \tag{3.2}$$

$$\text{(iii)} \qquad h(t) = \begin{cases} 0 & t \in [0, \alpha) \\ t & t \in [\alpha, \beta) \\ 1 & t \in [\beta, 1]. \end{cases}$$

Functionals of type (3.1) are mentioned within the theory of nonlinear utility (Yaari, 1987; Quiggin, 1982; Segal, 1984), but with a function $h(\cdot)$ which always turns out to be continuous, and thus inadequate for expressing the statistics mentioned above.

If F has support contained in $[a, b]$ and $h(0) = 0$ and $h(1) = 1$, letting $g(t) = 1 - h(t)$, some easy manipulations give

$$\int_a^b x dh(F(x)) = \int_a^b (1 - h(F(x)))dx + a = \int_a^b g(F(x))dx + a. \tag{3.3}$$

The RHS of (3.3) can be replaced by a more general expression, where no kind of continuity is required for $g(\cdot)$. We now proceed to give a characterization theorem for functionals which are continuous transforms of $\int g(F(x))dx$ with monotone $g(\cdot)$. These can be referred to as L-functionals, by analogy to L-statistics when $F(x)$ is the empirical distribution. In this theorem associativity is taken "laid on its side".

Let us extend the definition of F^{-1} for all $F \in D[a, b]$ by $F^{-1}(1) = b$; this is different from the quantile function for which $F^{-1}(1) = \sup \operatorname{supp}(F)$. The function $F^{-1} : [0, 1] \to [a, b]$ is nondecreasing and right continuous; observe that $(F^{-1})^{-1} = F$.

THEOREM 2. *A functional $\Phi : D[a, b] \to R$ satisfies*

(1*) $\Phi(F_1) = \Phi(F_2) \Rightarrow \Phi(((1 - \lambda)F_1^{-1} + \lambda G^{-1})^{-1}) = \Phi(((1 - \lambda)F_2^{-1} + \lambda G^{-1})^{-1})\ \forall\ F_1, F_2, G \in D[a, b]$, and $0 \leq \lambda \leq 1$,

(2*) $\Phi(\delta_x) = \phi(x)$ *is continuous in* $[a, b]$,

(3*) Φ *is increasing (not necessarily strictly) with respect to the stochastic ordering* $\leq^{st}$,

iff $\exists$ *two real functions* g *and* γ, *with* γ *continuous and strictly increasing and* $g(\cdot)$ *decreasing, such that*

$$\forall\ F \in D \quad \Phi(F) = \gamma\left(\int_a^b g(F(x))dx\right);$$

$g(\cdot)$ *is unique up to positive affine transformations. Note that the same is true for* Φ *decreasing in (3*) with an increasing* $g(\cdot)$.

PROOF. We can map $[a, b]$ by a linear transformation onto $[0, 1]$ and let $\widetilde{F}$ be the induced distribution function with support contained in $[0, 1]$. For the functional $\widetilde{\Phi}$ defined by $\widetilde{\Phi}(\widetilde{F}) = \Phi(F)$, properties (1*), (2*), (3*) are still true. Thus assume w.l.o.g. that $\Phi : D[0, 1] \to R$. Observe that $\widetilde{F}^{-1}$ can be thought of as a distribution function on $[0, 1]$; let $\Psi : D[0, 1] \to R$ be defined as $\Psi(F^{-1}) = \Phi(F)$. ∎

It is a straightforward check that (1*) $\Rightarrow$ (1) of Theorem 1 for Ψ. Also since $F \leq^{st} G \Rightarrow G^{-1} \leq^{st} F^{-1}$, by (3*) Ψ is monotone decreasing, since Φ is increasing, w.r.t. $\leq^{st}$. It is obvious that (2*) for $\Phi \Rightarrow$ continuity of the function $\Psi(\lambda)$ relative to Ψ; in particular if Φ is consistent, namely $\Phi(\delta_x) = x \ \forall \ x \in [a,b], \psi(\lambda) = 1-\lambda$. The conclusions of Theorem 1 can be applied to Ψ, namely there exist g, γ s.t. $\Phi(F) = \Psi(F^{-1}) = \gamma \int g(t)dF^{-1}(t) = \gamma \left(\int g(F)dx\right)$, and by the same theorem $g(\cdot)$ is unique up to positive affine transformations. The "if" part of the theorem is straightforward using the identity

$$\frac{1}{(b-a)} \int_a^b g(F(x))dx = \int_0^1 g(t)dF^{-1}(t). \qquad \blacksquare$$

In Theorem 2, the function $\phi(x) = \Phi(\delta_x)$ is monotonic by (3*). Condition (1*) ensures that it is 1-1 in (a,b), by Lemma 2. Replacing (2*) by "$\phi(x)$ is invertible", which is thus equivalent to "$\phi(x)$ is onto the whole range of Φ", would lead to the same result, by (2) of Lemma 3 (when $n = 1$) applied, as above, in a "dual" way. However such a condition does not appear to be intuitively very appealing, and is itself equivalent to the continuity of ϕ when range (Φ) is connected.

Since we are still dealing with measures of location, (3*) is a natural order replacement; (2*) is also natural, although the continuity condition is perhaps too general; if we want to represent "means" we would normally assume consistency. Condition (1*) says that we are taking mixtures "horizontally" rather than "vertically", namely mixtures of the actual ordered values of the r.v. rather than probability mixtures. From a mathematical point of view it may be fairly easy to check, say graphically, whether or not this condition holds. However (1*) is amenable to an intuitive meaning only when there is a natural way of matching the ordered values of the random variables in question so that the ranking is preserved going from one to the other. An economic interpretation in terms of preferences similar to Yaari's (1987) can be given. Another example is the following: say that F_1 and F_2 express two different distributions of firms by number of employees, referred to a given total workforce, and say that F_1 and F_2 are equivalent from the point of view of our measure Φ. (1*) states that the two populations are equivalent if more personnel is distributed in the same way among firms so as to preserve their ranking as regards number of staff. We shall comment further on this property at the end of Section 4.

The differences between our Theorem 2 and Yaari's (1987) Theorem 1 are that our assumptions (1*) and (2*) are weaker than his – we do not require "dual independence" – and mainly, which we would like to underline, that there is no need to prove this result anew, since it is really a consequence of our Theorem 1, namely a modified dFNK. In addition, our theorem is capable of taking in applications other than utilities.

COROLLARY. *Replacing conditions (2*) of Theorem 2 with continuity of* Φ

w.r.t. weak convergence, the same result holds and moreover $g(\cdot)$ is continuous.

PROOF. Convergence in distribution of a sequence of r.v. X_n is equivalent to convergence in quantile, namely weak convergence of F_n^{-1}. Thus we apply the Corollary of Theorem 1 to Ψ. ∎

4. Representing Some Measures of Dispersion and of Inequality. There are many ways of partially ordering distributions according to their dispersion: see for instance Shaked (1985) for a comprehensive survey. An important stochastic ordering of dispersion is *dilation*, which we shall denote by $\leq_D$. This is defined by the relation: let $F, G \in D(X)$, $X \subseteq R$ connected, then $F \leq_D G \Leftrightarrow \int_X f dF \leq \int_X f dG$ $\forall$ convex function f such that the integral exists, which roughly speaking means that G has "more variability" than F.

It can be shown that $F \leq_D G$ if and only if

$$\int_{-\infty}^{v} (F(x) - G(x))dx \leq 0 \qquad \text{for all } v \text{ in } X, v > -\infty$$

and

$$\int_X (F(x) - G(x))dx = 0. \tag{4.1}$$

Thus $F \leq_D G$ implies that F and G have the same expected value; for r.v.'s with the same expectation $\leq_D$ can be thought of as the reverse ordering to second degree stochastic dominance and coincides with the "spread" ordering of Bickel and Lehmann (1979). For frequency distributions from finite populations and for sample distributions $\leq_D$ coincides with *majorization* of n-tuples of real numbers and the order preserving functions w.r.t. such orderings are called *Schur convex* (Marshall and Olkin, 1979).

LEMMA 4. *If F, G are distributions with bounded support and the same expectation μ, an $G(x) - F(x)$ has at most one sign change (from positive to negative) as x ranges over $X = supp(F) \cup supp(G)$, then $F \leq_D G$.*

PROOF. See for instance Shaked (1980).

THEOREM 3. *Let $X \subseteq R$ be connected, and let $D_\mu(X)^B$ = the set of distributions with support bounded in X and given expectation $= \mu$; then Φ: $D_\mu(X)^B \to R$ is*

(1) *associative,*

(2) *continuous w.r.t. weights,*

(3) *increasing (decreasing) w.r.t. the dilation ordering;*

iff $\exists\ \ell : X \to R$ *convex (concave) and a real function* ψ, *continuous and increasing, such that*

$$\Phi(F) = \psi\left(\int_X \ell dF\right); \tag{4.2}$$

$\ell(\cdot)$ *is unique up to positive affine transformations.*

PROOF. This theorem needs a different proof from Theorem 1, since concentrated distributions other than δ_μ do not belong to $D_\mu(X)^B$. First of all we shall prove the result for $D_\mu[a,b]^0 =$ all distributions with finite support in a given interval $[a,b] \subseteq X$, then for all distributions with finite support in X and eventually for all distributions with bounded support. We write D instead of $D_\mu(X)^B$.

Let $a \le \mu \le b$, $a, b \in X$ fixed; let E be the set of distributions of $D_\mu[a,b]^0$ with support in at most 2 points. These can be written as:

if $a \le x \le \mu$ $\quad \Delta_{x,b} = (1-\beta(x))\delta_x + \beta(x)\delta_b \quad$ where $\beta(x) = (\mu - x)/(b-x)$
if $\mu \le x \le b$ $\quad \Delta_{a,x} = \alpha(x)\delta_a + (1-\alpha(x))\delta_x \quad$ where $\alpha(x) = (x-\mu)/(x-a)$

Observe that $\alpha(\mu) = \beta(\mu) = 0$ and $\Delta_{a,\mu} = \Delta_{\mu,b} = \delta_\mu$, and furthermore that $\beta(a) = 1 - \alpha(b)$ so that the symbol $\Delta_{a,b}$ is uniquely defined. Let $D^0_{a,b}$ denote all distributions in D with finite support and such that $a =$ minimum of the support and $b =$ maximum of the support, i.e. $D^0_{a,b} = \{F \in D : F = \lambda_o\delta_a + \Sigma\lambda_i\delta_{x_i} + \lambda_n\delta_b$ with $\lambda_o \ne 0$, $\lambda_n \ne 0$ and $a < x_i < b$, $i = 1, \cdots, n-1)$. We show first that $D^0_{a,b} \subseteq H(E)$. For simplicity the proof is given for distributions with support in three points only: it can be extended by induction to any number of points. So let

$$F = \lambda_o\delta_a + \lambda_1\delta_x + \lambda_2\delta_b,\ \ \lambda_o + \lambda_1 + \lambda_2 = 1,\ \ \lambda_i > 0,\ \ \lambda_o a + \lambda_1 x + \lambda_2 b = \mu \tag{4.3}$$

with, say, $\mu < x < b$; then

$$F = (1-\gamma)\Delta_{a,b} + \gamma\Delta_{a,x}$$

with $\gamma = \lambda_1/(1-\alpha(x)) = [\lambda_o - \alpha(b)]/[\alpha(x) - \alpha(b)] = 1 - \lambda_2/\beta(a)$. Moreover $D^0_{a,b}$ is "dense w.r.t. variation of weights" in $D_\mu[a,b]^0 = \{F \in D : F = \Sigma_i\lambda_i\delta_{x_i},\ x_i \in [a,b]\}$, in the sense that $F = \lim_{\eta\to 0} F_\eta$, $F_\eta \in D^0_{a,b}$, if we put $F_\eta = \eta\Delta_{a,b} + (1-\eta)F$.

Note further that $\forall\ F \in D_\mu[a,b]\ \delta_\mu \le_D F \le_D \Delta_{a,b}$, so $\Phi(\delta_\mu) = \min\{\Phi(F) : F \in D_\mu[a,b]\}$, $\Phi(\Delta_{a,b}) = \max\{\Phi(F) : F \in D_\mu[a,b]\}$. By Lemma 3 the function $\psi(\lambda) = \Phi((1-\lambda)\delta_\mu + \lambda\Delta_{a,b})$ is continuous, increasing, invertible and such that $\psi^{-1}\Phi$ is linear on convex combinations. Now define $\ell(x)$ in $[a,b]$ by means of

$$\text{if } a \le x \le \mu \qquad \psi^{-1}\Phi(\Delta_{x,b}) = (1-\beta(x))\ell(x) + \beta(x)\ell(b)$$

where the constant $\ell(b)$ is chosen arbitrarily, and

$$\text{if } \mu \leq x \leq b \qquad \psi^{-1}\Phi(\Delta_{a,x}) = \alpha(x)\ell(a) + (1-\alpha(x))\ell(x)$$

where $\ell(a)$ is derived from the previous expression. Note that $\ell(\mu) = 0$. We show that ℓ satisfies the representation: for all distributions in $D^0_{a,b}$, e.g. for F as in (4.3), a simple calculation will suffice, using linearity of $\psi^{-1}\Phi$, and the same is true in all $D_\mu[a,b]^0$ by continuity of Φ w.r.t. weights. Thus for this function $\ell(\cdot)$, $\Phi(\Sigma_i\lambda_i\delta_{x_i}) = \psi(\Sigma_i\lambda_i\ell(x_i))$.

We now show that ℓ is convex. Let $X_1, X_2 \in [a,b]$ and put $\bar{x} = (1-\lambda)x_1 + \lambda x_2$; say $\bar{x} \leq \mu$. Define $F = \lambda_1\delta_{x_1} + \lambda_2\delta_{x_2} + \lambda_3\delta_b$ with $\lambda_1 = (1-\lambda)(1-\beta(\bar{x}))$, $\lambda_2 = \lambda(1-\beta(\bar{x}))$, $\lambda_3 = \beta(\bar{x})$, so that $F \in D_\mu[a,b]^0$. It is easy to see by Lemma 4 that $\Delta_{\bar{x},b} \leq_D F$. Thus $\Phi(\Delta_{\bar{x},b}) \leq \Phi(F)$, whence

$$\begin{aligned}\psi^{-1}\Phi(\Delta_{\bar{x},b}) &= (1-\beta(\bar{x}))\ell(\bar{x}) + \beta(\bar{x})\ell(b)\\ &\leq \psi^{-1}\Phi(F) = (1-\lambda)(1-\beta(\bar{x}))\ell(x_1)\\ &\quad + \lambda(1-\beta(\bar{x}))\ell(x_2) + \beta(\bar{x})\ell(b),\end{aligned}$$

and the convexity of ℓ follows. This shows that the assertion holds for distributions with finite support in $[a,b]$. A classical argument can be employed to prove uniqueness of $\ell(\cdot)$ up to affine transformations.

In order to extend the representation to X, take an increasing sequence of intervals $[a_n, b_n]$, $a_n \leq \mu \leq b_n$, with $\lim a_n = \inf X$, $\lim b_n = \sup X$. By Lemma 3 there exists ψ on X, continuous and strictly increasing, such that $\psi^{-1}\Phi$ is linear on convex combinations. Define $\ell_n(x)$ in each $[a_n, b_n]$ as before. Such an ℓ_n makes the theorem hold true for any F in $D[a_n, b_n]^0$. Choosing for every n an affine transformation appropriately, i.e. in such a way that $\ell_n(b_1) = \ell_1(b_1)$ and $\ell_n(\mu) = 0$, one gets $\ell_n(x) = \ell_{n-1}(x)$ for all $x \in [a_{n-1}, b_{n-1}]$. Thus it is possible to define a convex $\ell : X \to R$ by $\ell(x) = \ell_n(x)$ for $x \in [a_n, b_n]$.

Lastly, let $F \in D$ have bounded support. The difficulty lies in finding two sequences $\{F_n\}$ and $\{G_n\}$ of distributions with finite support such that $\forall\ n\ E(F_n) = E(G_n) = \mu$ and $F_n \leq_D F \leq_D G_n$, and further if $n > n_\epsilon$ $\Phi(G_n) - \Phi(F_n) = \int \ell dG_n - \int \ell dF_n < \epsilon$. If so, $\Phi(F_n) \leq \Phi(F) \leq \Phi(G_n)$ by assumption (3) and $\int \ell dF_n \leq \int \ell dF \leq \int \ell dG_n$ since ℓ is convex: the conclusion follows from $\int \ell dF_n = \Phi(F_n)$, $\int \ell dF_n = \Phi(G_n)$. Let $\widetilde{F}_n = \Sigma_i\lambda_i\delta_{x_i}$ $(i = 1, \cdots, n)$ and $\widetilde{G}_n = \Sigma_i\lambda_{i+1}\delta_{x_i}$, $(i = 0, 1, \cdots, n-1)$ where $F(x_0) = 0$, $F(x_n) = 1$ and $\lambda_i = F(x_i) - F(x_{i-1}) < \epsilon/V$, where V is the variation of $\ell(x)$ in the support of F. Clearly $E(\widetilde{F}_n) \geq \mu \geq E(\widetilde{G}_n)$ and equally clearly the difference of the integrals is $< \epsilon$. If $E(\widetilde{F}_n) \neq \mu$, there exists a k and a point $x_k \leq y \leq x_{k+1}$ such that $F_n = \sum_{i=1}^{k}\lambda_i\delta_{x_i} + \lambda_{k+1}\delta_y + \sum_{i=k+1}^{n-1}\lambda_{i+1}\delta_{x_i}$ has $E(F_n) = \mu$. As the graph of F_n crosses that of F just once (in y), it follows from Lemma 4 that $F_n \leq_D F$. A symmetric argument leads to $F \leq_D G_n$. For these distributions the theorem is true.

The converse follows immediately from the properties of ψ and from the linearity of the integral with respect to F. ∎

Theorem 3 can be applied within the framework of measures of the dispersion of a distribution about a given expected value. The theorem characterizes the ones that are associative, "robust" and preserve the dilation order.

A different characterization was given by Peccati and Regazzini (1977): choosing a measure of deviation $d(x, \theta)$ of x from a "center" θ and applying dFNK to the distribution of $|d(x, \theta)|$ they obtain all dispersion measures of the form

$$u^{-1} \int u(|d(x, \theta)|) dF, \tag{4.4}$$

with $u(\cdot)$ increasing and continuous. Some well known indices of dispersion, like the standard deviation and all absolute central moments, can be seen to satisfy both (4.2) and (4.4). However, even when $\theta = \mu$ the two families described by (4.2) and (4.4) are not the same, and in any case the main difference between their result and our Theorem 3 lies in the initial assumptions.

By interpreting the function $\ell(x)$ as a loss function, Theorem 3 also characterizes a class of *risk functions* with respect to convex losses.

A dual result is

THEOREM 4. *Let $D_\mu[0,1]$ be the distributions with support in $[0,1]$ and given expectation $= \mu$; then $\Phi : D_\mu[0,1] \to R$ satisfies*

(1*) $\Phi(F_1) = \Phi(F_2) \Rightarrow \Phi((\lambda F_1^{-1} + (1-\lambda)G^{-1})^{-1}) = \Phi((\lambda F_2^{-1}(1-\lambda)G^{-1})^{-1}) \; \forall \, F_1, F_2, G \in D, \forall \, 0 \leq \lambda \leq 1$.

(2*) $\Phi(\Sigma_i \lambda_i \delta_{x_i})$ *is continuous w.r.t. the* x_i;

(3*) Φ *is increasing (not strictly) with respect to the Lorenz ordering,*

iff $\exists \, u : [0,1] \to [0,1]$ *concave and a real function ψ continuous and increasing such that*

$$\Phi(F) = \psi\left(\int u(F) dx\right). \tag{4.5}$$

PROOF. We recall the definition of *Lorenz ordering* for nonnegative random variables with finite expectation (see also Arnold, 1986):

$$X \leq_L Y \Leftrightarrow L_X(v) \geq L_Y(v) \quad \text{for all } v \in [0,1],$$

where $L_X(v)$ is the *Lorenz curve* defined as

$$L_X(v) = \left(\int_0^v F^{-1}(t)dt\right) / \left(\int_0^1 F^{-1}(t)dt\right) = \mu^{-1}\left(\int_0^v F^{-1}(t)dt\right).$$

If F, $G \in D_\mu[0,1] \Rightarrow F^{-1}$, G^{-1} have the same expectation. And in this case the Lorenz ordering for the F's is the reverse of the dilation ordering for their inverses. Also (2*) implies continuity w.r.t. weights for the functional $\Psi(F^{-1}) = \Phi(F)$. Thus Ψ satisfies the assumptions of Theorem 3, with Ψ decreasing in (3), hence the integral representation holds with a concave function $u(\cdot)$. ∎

Inequality, i.e. the difference between the actual frequency distribution of a character in a population and the distribution one would get if that character was possessed equally by all the units of that collective, is usually compared through the Lorenz order, and traditionally refers to distribution of income. Observe that for bounded nonnegative r.v.'s with the same expectation $\leq_L$ coincides with $\leq_D$ (Arnold, 1986, Theorem 3.2). Thus Theorem 3 states that inequality measures (also called concentration indices) are increasing continuous transforms of the expected value of a convex function if and only if they are associative (and mildly) robust (in the sense of condition (2)). Examples of such inequality measures are Pietra's index: $\frac{1}{2}\mu^{-1}\Sigma_i p_i|x_i - \mu|$, Theil's: $\Sigma_i p_i(x_i/\mu)\log(x_i/\mu)$ and many more.

Theorem 4 characterizes those concentration indices that satisfy "dual" associativity, i.e. (1*), and a different type of mild robustness, given by (2*). We can interpret (1*) as follows: if the same total income is distributed in two populations so that the concentration is the same w.r.t. Φ, (1*) states that the concentration is still the same if in addition a further income is distributed so that "the poor get less and the rich get more," i.e. the poor get smaller proportions and the rich get larger proportions in such a way as to preserve the ranking. An example is Gini's measure of income inequality which can be expressed as $\mu^{-1}\int F(x)(1 - F(x))dx$, namely is of the form (4.5); this is continuous but not associative: it is however associative in the dual sense.

Acknowledgments. We wish to thank two referees for helpful comments. This paper was written while the authors were participants of two research projects sponsored by MURST (Italian Ministry for University and Scientific and Technological Research), namely "Orderings for Stochastic Processes and Applications" and "Descriptive Analyses and Interpretation of Data."

REFERENCES

ARNOLD, B. C. (1986). *Majorization and the Lorenz Order: A Brief Introduction.* Springer–Verlag, New York.

BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models I: Introduction. *Ann. Statist.* **3**, 1038–1044.

BICKEL, P. J. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models IV: Spread. In *Contributions to Statistics, J. Hajek Memorial Volume.* (J. Jureckova, ed.) Reidel, Dordrecht, 33–40.

CHEW, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica* **51**, 1065–1092.

CHEW, S. H. (1989). Axiomatic utility theories with the betweenness property. *Ann. Oper. Res.* **19**, 273–298.

DABONI, L. and WEDLIN, A. (1982). *Statistica: un'Introduzione all'Impostazione Neo-Bayesiana.* U.T.E.T., Torino.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

DE FINETTI, B. (1931). Sul concetto di media. *Giorn. Ist. Ital. Attuari.* **2**, 369–396.

FISHBURN, P. C. (1986). Implicit mean value and certainty equivalence. *Econometrica* **54**, 1197–1205.

FISHBURN, P. C. and VICKSON, R. G. (1978). Theoretical foundations of stochastic dominance. In *Stochastic Dominance.* (G. H. Whitmore and M. C. Findlay, eds.) Lexington Books, Heath, MA.

GIOVAGNOLI, A. and REGOLI, G. (1989). Generalizations of the de Finetti-Nagumo-Kolmogorov theorem. Technical Report No. 3/1989, Department of Mathematics, University of Perugia.

GIOVAGNOLI, A. and REGOLI, G. (1990) Generalizzazioni del teorema di de Finetti-Nagumo-Kolmogorov. *Atti XXXV Riunione Scientifica S.I.S.*, Padova 18–21 Aprile 1990, Vol. 2, 129–136.

HARDY, G. H., LITTLEWOOD, J. E. and POLYA, G. (1984). *Inequalities.* Cambridge University Press, New York.

HUBER, P. (1981). *Robust Statistics.* John Wiley & Sons, New York.

KOLMOGOROV, A. N. (1930). Sur la notion de la moyenne. *Atti Rend. Accad. Lincei* **12**, 388–391.

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

NAGUMO, M. (1930). Über eine klasse der Mittelwerte. *Japan J. Math.* **7**, 71–79.

PECCATI, L. and REGAZZINI, E. (1977). Sugli indici di variabilità. *Giorn. Econom. Ann. Econom.*, 145–167.

QUIGGIN, J. (1982). A theory of anticipated utility. *J. Econom. Behaviour Organization* **3**, 225–243.

SEGAL, U. (1984). Nonlinear decision weights with the independence axiom. University of California, Los Angeles, Working Paper 353.

SHAKED, M. (1985). Ordering distributions by dispersion. In *Encyclopedia of Statistical Sciences,* **6**. (N. L. Johnson and S. Kotz, eds.) John Wiley & Sons, New York, 485–490.

SHAKED, M. (1980). On mixtures from exponential families. *J. Roy. Statist. Soc. Ser. B* 192–198.

YAARI, M. E. (1987). The dual theory of choice under risk. *Econometrica* **55**, 95–115.

DIPARTIMENTO DI SCIENZE STATISTICHE
UNIVERSITÀ DEGLI STUDI DI PERUGIA
VIA A. PASCOLI - C.P. 1315/1
06100 PERUGIA, ITALY

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DI PERUGIA
VIA VANVITELLI 1
06100 PERUGIA, ITALY