The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 466–480

CONCENTRATION OF RANDOM QUOTIENTS[1]

BY WILLIAM H. LAWTON

Eastman Kodak Company

1. Summary. The present paper proposes a definition of relative concentration of random variables about a given constant, and studies the relationship between two stochastic denominators Z_1 and Z_2 which causes the random quotient X/Z_1 to be more concentrated about zero than X/Z_2. In this paper we shall always assume that the numerator and denominator are independent.

Two necessary and sufficient conditions, and several sufficient conditions for X/Z_1 to be more concentrated about zero than X/Z_2 are given in Section 4. The results of Section 4 are used in Section 5 to obtain generalizations of a theorem due to Hájek (1957) on the generalized Student's t-distribution. Sections 6 and 7 use these generalized theorems to produce tests and confidence intervals for several Behrens-Fisher type problems. Finally, Section 8 contains a proof of the randomization theorem stated in Section 4.

In particular, Secton 6 is concerned with the extension of a result in Lawton (1965) on Lord's u-statistic to the case of unequal sample sizes. Section 7 gives methods for constructing confidence intervals for linear combinations of means from k normal populations.

2. Introduction. While the definition of relative concentration of random variables considered in this paper has not been given previously, the concept is quite closely related to the notion of relative peakedness presented by Birnbaum (1948), and has appeared, somewhat disguised, in a paper by Hájek (1957).

DEFINITION 1. A random variable Z is *more concentrated about a constant* c_0 than the random variable Y, in symbols $Z \ll_{c_0} Y$, provided

$$P\{a < Z - c_0 < b\} \geqq P\{a < Y - c_0 < b\}$$

for every interval (a, b) containing zero.

Let Z_a be a random variable with distribution function $F((z - b)/a)$, that is, a random variable with location parameter b and scale parameter a, then Z_{a_1} is more concentrated about b than Z_{a_2} if and only if $a_1 \leqq a_2$. Thus, the notion of concentration given by Definition 1 is consistent with the notion of concentration in terms of scale parameter. Also, as one would hope, the random variable degenerate at c_0 is the most concentrated random variable about c_0. It may be of interest to note that concentration about c_0 has a simple interpretation in terms of distribution functions.

Received 23 February 1967; revised 18 October 1967.

[1] This paper was prepared with the partial support of the United States Army under contract No. ARO-D-31-124-D548, while the author was at the University of California, Berkeley.

REMARK 1. $Z \ll_{c_0} Y$ *if and only if*

$$F_Z(x) \leqq F_Y(x) \quad \text{for} \quad x < c_0,$$
$$F_Z(x) \geqq F_Y(x) \quad \text{for} \quad x > c_0.$$

The proof is immediate.

If the distributions of Y and Z in Definition 1 are symmetric about c_0, as they were in Birnbaum's work, then the present notion of concentration coincides with Birnbaum's concept of relative peakedness. Birnbaum was concerned with concentration of sums (or equivalently, of arithmetic means) of random variables, and he proved

THEOREM 1. (Birnbaum). *Let Y_1, Y_2, Z_1, Z_2 be continuous random variables with probability densities φ_1, φ_2, f_1, f_2 respectively. If*

(i) *Y_1 and Y_2 are independent, and Z_1 and Z_2 are independent,*

(ii) *φ_i and f_i, $i = 1, 2$, are symmetric about zero,*

(iii) *φ_2 and f_1 are non-increasing functions on $[0, \infty)$,*

then

$$Y_i \ll_0 Z_i, \; i = 1, 2 \Rightarrow Y_1 + Y_2 \ll_0 Z_1 + Z_2.$$

This result relates concentration of individual variables to concentration of their sums. From this theorem he obtained

THEOREM 2. (Birnbaum). *Let Y be a continuous random variable such that*

(i) *$\varphi(y)$ is symmetric about zero,*

(ii) *$\varphi(y)$ is non-increasing on $[0, \infty)$,*

(iii) *$P\{|Y| > a\} = 0$ for some finite $a > 0$,*

then for all n

$$W_0 \ll_0 \bar{Y}_n \ll_0 \bar{Z}_n$$

where $\bar{Y}_n$ is the sample mean of n independent observations of Y, and $\bar{Z}_n$ is the sample mean of n independent observations from the uniform distribution on $(-a, a)$, and W_0 the random variable degenerate at 0.

Now concentration about 0 ($\ll_0$) is an order relation on the space of random variables, and this last theorem merely gives the extreme points (relative to $\ll_0$) of a special set of random variables. If $\mathcal{Y}$ is the class of all random variables satisfying conditions (i), (ii), and (iii) of Theorem 2, then this theorem states that the set $\{\bar{Y}_n : Y_i$ independent, identically distributed, and in $\mathcal{Y}\}$ has extreme points W_0 and $\bar{Z}_n$. Birnbaum uses the bound $\bar{Z}_n$ to obtain an upper bound for the tail probabilities $P\{|\bar{Y}_n| > y\}$.

Hájek (1957) does not mention the notion of concentration or peakedness, but his results are easily expressed in terms of these concepts. Unlike Birnbaum, Hájek dealt with ratios of independent random variables.

THEOREM 3. (Hájek). *Let X have a standard normal distribution, and let $\mathcal{Z} = \{Z : Z = \sum_{i=1}^{n} \lambda_i \chi_i^2(1), \lambda_i \geqq 0, \sum_{i=1}^{n} \lambda_i = 1$ where the $\chi_i^2(1)$ are independent of each other, as well as of $X\}$. Then*

(1) $$X/[\chi^2(n)/n]^{\frac{1}{2}} \ll_0 X/[Z]^{\frac{1}{2}} \ll_0 X/[\chi^2(\nu)/\nu]^{\frac{1}{2}}$$

for all Z in $\mathcal{Z}$, *provided* $\nu \leqq \min_{1 \leqq i \leqq n} \{1/\lambda_i\}$.

Thus, Hájek examined the class of random quotients $\{X/[Z]^{\frac{1}{2}}: X$ distributed $N(0, 1), Z \varepsilon \mathcal{Z}\}$, and demonstrated the existence of extreme points (relative to $\ll_0$) in the set. Thus, Hájek, like Birnbaum, has examined a set of random variables and shown that the set has extreme points relative to the ordering $\ll_0$. Hájek used his result to obtain bounds for the type I error in the Behrens-Fisher problem.

One of the questions immediately raised by Hájek's work is whether one need be so restrictive about the distributions of the numerator and denominator, or do similar results about extreme points hold for more general classes of random quotients. Also, one might ask what property of the denominator causes $X/Z_1 \ll_0 X/Z_2$. Clearly, if Z_1 and Z_2 were non-negative and Z_1 were stochastically larger than Z_2, then this order relation would hold. This, however, is not a necessary condition because $[\chi^2(n)/n]^{\frac{1}{2}}$ is not stochastically larger than $[\chi^2(\nu)/\nu]^{\frac{1}{2}}$.

3. Assumptions, remarks, and notation. As in [3] we will consider random quotients of the form $X/h(Z)$, where X and Z are independent random variables, Z non-negative and $h(\cdot)$ a continuous real-valued function. In Hájek's work X is standard normal, $h(z) = z^{\frac{1}{2}}$, $Z = \sum_{i=1}^{n} \lambda_i \chi_i^2(1)$. It was shown in [3] that Hájek's proof goes through if one only assumes that the following condition is satisfied.

CONDITION (A). *The pair* (X, h) *is said to satisfy condition* (A) *if*

(1) X *has a continuous distribution function,*

(2) *for any* $a < 0, b > 0$ *the function* $f_{a,b}(z) = P\{a < X/h(z) < b\}$ *is concave for* $z \geqq 0$.

EXAMPLE 1. If $h(z) = z$, and X is a random variable with a density $\varphi(x)$, then Condition (A) simply requires that $\varphi(x)$ be unimodal at zero. That is, $\varphi(x)$ and $\varphi(-x)$ are non-increasing functions on $[0, \infty)$. This then would allow X to be distributed $N(\delta, 1)$ only when $\delta = 0$.

EXAMPLE 2. Once again let X have a density $\varphi(x)$. If we allow $h(\cdot)$ to be a concave function on $[0, \infty)$, then we can relax the restriction that $\varphi(x)$ be unimodal at zero. For example, let X be distributed $N(\delta, 1)$ and $h(z) = z^{\frac{1}{2}}$, then it can be shown that the pair (X, h) satisfies Condition (A) if and only if $|\delta| \leqq 2$.

The relationship (1) of Hájek's Theorem 3 states that for $n \geqq \nu$

(2) $$X/h(\chi^2(n)/n) \ll_0 X/h(\chi^2(\nu)/\nu)$$

whenever the numerators and denominators are independent, and X has a standard normal distribution and $h(z) = z^{\frac{1}{2}}$. However, since the proof Hájek used works for any (X, h) satisfying Condition (A), we find that the two non-negative random variables $\chi^2(n)/n$ and $\chi^2(\nu)/\nu$ have some relationship which causes (2) to hold whenever the numerators and denominators are independent and (X, h) satisfies Condition (A). The following example shows, however, that (2) may not hold without Condition (A).

EXAMPLE 3. Let X be distributed $N(\delta, 1)$ with $\delta = 4.25$, and let $h(z) = z^{\frac{1}{2}}$, then it can be shown that (X, h) violates the second requirement of Condition (A). Take $n = 10$ and $\nu = 5$, then (2) is equivalent to

$$P\{a < t_{10}(4.25) < b\} \geqq P\{a < t_5(4.25) < b\}$$

for all intervals (a, b) containing zero. Here $t_n(\delta)$ represents a random variable having the non-central t-distribution with n degrees of freedom and non-centrality parameter δ. Tables of the non-central t-distribution yield

$$P\{t_{10}(4.25) < 6.0\} = .7437 < .8210 = P\{t_5(4.25) < 6.0\}$$

which contradicts the relationship in (2).

The relationship in (2) leads us to define the notion of a uniformly better denominator.

DEFINITION 2. Let Z_1 and Z_2 be two non-negative random variables. Z_1 is said to be a *uniformly better denominator* than Z_2, written $Z_1 \prec Z_2$, provided

$$X/h(Z_1) \ll_0 X/h(Z_2)$$

whenever X is independent of Z_1 and Z_2, and (X, h) satisfies Condition (A).

Our goal in the next section is the study of the order relation of Definition 2. In the remainder of the paper let $\bar{R}^+$ denote the compact, positive half-line.

4. Properties and characterization of $\prec$. We now proceed to investigate those properties of Z_1 and Z_2 which cause $Z_1 \prec Z_2$. Theorem 4 gives two necessary and sufficient conditions for $Z_1 \prec Z_2$ which will be useful in studying the ordering. Theorem 4(iii) is of particular interest because it describes a method by which one may construct, for a given Z_1, all random variables Z such that $Z_1 \prec Z$.

Let $\mathcal{S}$ represent the class of all non-negative, concave, non-decreasing continuous functions on $\bar{R}^+$ to $\bar{R}^+$. Define the notion of a randomized, mean decreasing map, rmd map for short, as follows:

DEFINITION 3. Any real valued function $T_z(y)$ on $\bar{R}^+ \times \bar{R}^+$ which is

(1) a measurable function in z for each fixed y,

(2) the distribution function of a non-negative random variable for each fixed z, and

(3) $\int_{\bar{R}^+} y\, dT_z(y) \leqq z$, for every z in $\bar{R}^+$,

is called a *randomized, mean decreasing map.*

For each non-negative random variable Z we will designate by $T(Z)$ the random variable having distribution function

$$F_{T(Z)}(y) = \int_{\bar{R}^+} T_z(y)\, dF_Z(z).$$

EXAMPLE 4. Let Z be any random variable defined on the non-negative integers, and let Y be a random variable with conditional distribution $P\{Y = k \mid Z = z\} = \binom{z}{k}p^k(1 - p)^{z-k}$ for some $0 \leqq p \leqq 1$. We have $E(Y \mid Z = z) = pz \leqq z$, and so

$$T_z(y) = P\{Y < y \mid Z = z\} = \textstyle\sum_{k<y} \binom{z}{k}p^k(1 - p)^{z-k}$$

is an rmd map. If Z has Poisson distribution $P\{Z = z\} = e^{-\lambda}\lambda^z/z!$, then $T(Z)$ is just the random variable Y whose distribution is given by

$$P\{Y = k\} = \sum_{z=0}^{\infty} P\{Y\, k \mid Z = z\}P\{Z = z\} = e^{-p\lambda}(p\lambda)^k/k!.$$

Thus, $T(Z)$ has Poisson distribution with parameter $p\lambda$.

The random variable degenerate at zero is clearly the worst possible denominator, since it can be obtained from any non-negative random variable by the rmd map

$$\begin{aligned} T_z(y) &= 1 \quad \text{for} \quad y < 0, \\ &= 0 \quad \text{for} \quad y \geqq 0. \end{aligned}$$

We are now in a position to state

THEOREM 4. *For any non-negative random variables Z_1 and Z_2, the following three statements are equivalent*:

(i) $Z_1 \prec Z_2$,

(ii) $Ef(Z_1) \geqq Ef(Z_2)$ *for all* $f \in \mathcal{S}$,

(iii) $Z_2 = T(Z_1)$ *where* T *is an* rmd *map.*

PROOF. First note that whenever (X, h) satisfies Condition (A) with X independent of Z we have

$$\begin{aligned} P\{a < X/h(Z) < b\} &= EP\{a < X/h(Z) < b \mid Z\} \\ &= Ef_{a,b}(Z) \end{aligned}$$

where $f_{a,b}(z)$ is concave, non-negative, non-decreasing, and continuous on $\bar{R}^+$ for any $a < 0$, $b > 0$. Thus, if (ii) of Theorem 4 holds, then for any interval (a, b) containing zero

$$\begin{aligned} P\{a < X/h(Z_1) < b\} &= Ef_{a,b}(Z_1) \\ &\geqq Ef_{a,b}(Z_2) = P\{a < X/h(Z_2) < b\} \end{aligned}$$

whenever X is independent of Z_1 and Z_2 and (X, h) satisfies Condition (A). From Definition 2 it then follows that (ii) implies (i).

Conversely, if $Z_1 \prec Z_2$ then it is clear that we have

$$Ef(Z_1) \geqq Ef(Z_2) \quad \text{for all } f \in \mathcal{S}' \tag{3}$$

where $\mathcal{S}'$ is the class of all concave, non-decreasing continuous functions on $\bar{R}^+$ with $f(0) = 0$ and $f(x) \to 1$ as $x \to \infty$. This inequality holds because any $f \in \mathcal{S}'$ is the distribution function of a non-negative random variable. Let X be independent of Z_1 and Z_2 and have $F_X(x) = f(x)$, and take $h(z) = z$. Under these conditions (X, h) satisfies Condition (A) and

$$Ef(Z_1) = P\{X/Z_1 < 1\} \geqq P\{X/Z_2 < 1\} = Ef(Z_2).$$

What remains then is to show that (3) holds for all f in $\mathcal{S}$. Since (3) holds for all f in $\mathcal{S}'$, it also holds for all non-negative, non-decreasing, concave, continuous functions on $\bar{R}^+$ bounded by $c < \infty$. This follows because any such bounded function can be transformed into a member of $\mathcal{S}'$ by subtraction of a suitable constant, and then division by some other constant. Now, for any f in $\mathcal{S}$, define

$$\begin{aligned} f_n(z) &= f(z) \quad \text{for } f(z) \leqq n, \\ &= n \qquad \text{otherwise.} \end{aligned}$$

We know that $Ef_n(Z_1) \geqq Ef_n(Z_2)$ for every n, and since $0 \leqq f_n(Z_i) \uparrow f(Z_i)$ we have, by the monotone convergence theorem, $Ef(Z_1) \geqq Ef(Z_2)$ for all f in $\mathcal{S}$. This proves the equivalence of (i) and (ii). The proof of equivalence of (ii) and (iii) will be deferred to Section 8. The equivalence follows at once from Theorem 10 of that section.

Let $M_Z(t)$ denote the moment generating function of the random variable $-Z$, where Z itself is a nonnegative random variable. That is, $M_Z(t) = Ee^{-tZ}$ for $t > 0$. For all $Z \geqq 0$, $M_Z(t)$ exists and is finite for $t > 0$. We have then

COROLLARY 1. *$Z_1 \prec Z_2$ implies that $M_{Z_1}(t) \leqq M_{Z_2}(t)$ for all $t > 0$.*

The proof follows immediately from the fact that $f_t(z) = 1 - e^{-tz}$ is in $\mathcal{S}$.

THEOREM 5. *Let Z_1 and Z_2 be non-negative random variables, then any of the following conditions are sufficient for $Z_1 \prec Z_2$:*

(i) *$Z_1 = E^{\mathcal{B}}Z_2$, where $\mathcal{B}$ is any σ-field contained in the σ-field generated by Z_2,*

(ii) *$Z_1 \ll_c Z_2$ and $EZ_2 \leqq EZ_1$,*[2]

(iii) *Z_1 is stochastically larger than Z_2.*

PROOF. The proof of (i) follows at once from Jensen's inequality and Theorem 4(ii), since for all concave functions one has $Ef(Z_1) = Ef(E^{\mathcal{B}}Z_2) \geqq E(E^{\mathcal{B}}f(Z_2)) = Ef(Z_2)$. The proof of (iii) is also simple. One again uses Theorem 4(ii) and the fact that $Ef(Z_1) \geqq Ef(Z_2)$ for any non-decreasing function f and Z_1 stochastically larger than Z_2.

The proof of part (ii) of the present theorem is based on a geometric interpretation of expectation. For any non-negative random variable Z with distribution function $F_Z(z)$, we have

$$EZ = \int_0^\infty [1 - F_Z(z)]\, dz.$$

That is, EZ is the area above the distribution function, and below the line $y = 1$.

Let us assume that $Z_1 \ll_c Z_2$ and $EZ_2 \leqq EZ_1$. Since $EZ_2 = EZ_1 + \int_c^\infty \delta(x)\, dx - \int_0^c \delta(x)\, dx$, we have $\int_0^c \delta(x)\, dx \geqq \int_c^\infty \delta(x)\, dx$ where $\delta(x)$ is defined as $|F_{Z_1}(x) - F_{Z_2}(x)|$. If A denotes the first of these integrals and B the second, we have $B \leqq A < \infty$, and both the areas are finite. Let $F_{s_1}(x)$ and $F_{s_2}(x)$ denote the distribution functions of the transformed random variables $s(Z_1)$ and $s(Z_2)$ where $s \in \mathcal{S}'$. The result will follow from Theorem 4(ii) if we can show $Es(Z_1) \geqq Es(Z_2)$ for all s in $\mathcal{S}'$ (the set of continuous, concave, non-decreasing functions on $\bar{R}^+$ with $s(0) = 0$ and $s(x) \rightarrow 1$ as $x \rightarrow \infty$). Because these $s(\cdot)$ are bounded by 1, both $Es(Z_1)$ and $Es(Z_2)$ are finite. Also $s(\cdot)$ is continuous and strictly

---

[2] Note: This condition on the expectations may be replaced by the weaker condition $\int_0^c \delta(x)\, dx \geqq \int_c^\infty \delta(x) dx$ where $\delta(x) = |F_{Z_1}(x) - F_{Z_2}(x)|$.

increasing on the interval $[0, a)$ defined by $\{x:s(x) < 1\}$; $s(\cdot)$ is constant for $x \geqq a$. For $0 \leqq x < a$, we can define $s^{-1}(\cdot)$. Thus, we have

$$F_{s(Z)}(y) = F_Z(s^{-1}(y)) \quad \text{for} \quad 0 \leqq y < 1,$$
$$= 1 \qquad\qquad \text{for} \quad y \geqq 1.$$

From this we can show that $s(Z_1)$ is more concentrated about some constant than $s(Z_2)$. First consider the case where $a > c$. We have $F_{s_1}(y) = F_{s_2}(y) = 1$ for $y \geqq 1$. For $0 \leqq y < 1$ we have

$$F_{s_i}(y) = F_{Z_i}(s^{-1}(y)) \qquad \text{for } i = 1, 2.$$

From this it follows that

$$F_{s_1}(y) = F_{Z_1}(s^{-1}(y)) \geqq F_{Z_2}(s^{-1}(y)) = F_{s_2}(y)$$

for $c \leqq s^{-1}(y) < a$; that is, $s(c) \leqq y < 1$. Similarly, we have

$$F_{s_1}(y) = F_{Z_1}(s^{-1}(y)) \leqq F_{Z_2}(s^{-1}(y)) = F_{s_2}(y)$$

for $0 \leqq y < s(c)$. Thus, we have shown that when $a > c$, $s(Z_1) \ll_{s(c)} s(Z_2)$. Because of this concentration relation we can define the transformed areas $A' = \int_0^{s(c)} \delta'(y)\,dy$, and $B' = \int_{s(c)}^1 \delta'(y)\,dy$ where $\delta'(y) = |F_{s_1}(y) - F_{s_2}(y)| = \delta(s^{-1}(y))$. Since $Es(Z_1)$ and $Es(Z_2)$ are finite, we have

$$Es(Z_1) = Es(Z_2) + [A' - B'],$$

and we need only show $A' \geqq B'$, for all s in $\mathcal{S}'$.

For $0 \leqq x < a$ we know $s(\cdot)$ is strictly increasing and continuous. The change of variable theorem for Riemann-Stieltjes integrals yields

$$A' = \int_0^{s(c)} \delta'(y)\,dy = \int_0^c \delta'(s(x))\,ds(x) = \int_0^c \delta(x)\,ds(x).$$

In our case $s(\cdot)$ is concave and thus we know the derivative exists everywhere, except possibly on a countable set. Let $s'(x)$ denote the left derivative, then

$$A' = \int_0^c \delta(x)s'(x)\,dx \geqq s'(c)\int_0^c \delta(x)\,dx = s'(c)A.$$

This last statement follows since, by concavity, $s'(x)$ is a non-increasing function of x. Using this same approach we can show $B' \leqq s'(c)B$. We then have $A' \geqq s'(c)A \geqq s'(c)B \geqq B'$ as was to be shown.

We now consider the case where $a \leqq c$. As before, $F_{s_1}(y) = F_{s_2}(y) = 1$ for $y \geqq 1$, and for $y < 1$ we have $F_{s_1}(y) \leqq F_{s_2}(y)$. Thus, in this case $s(Z_1) \ll_1 s(Z_2)$, and the transformed area $B' = 0$. Since $A' \geqq 0$ we again have $A' \geqq B'$. This completes the proof.

Theorem 5(ii) states the conditions under which greater concentration of the denominators implies greater concentration of the random quotients about zero. This particular result is very much like Theorem 1 of Birnbaum. Compare Theorem 1 with the following corollary.

COROLLARY 2. *Let X, Z_1, Z_2 be continuous random variables where X has* *density φ and the Z_i are non-negative. If*

(i) *Z_1 and Z_2 are independent of X,*

(ii) *$\varphi(x)$ and $\varphi(-x)$ are non-increasing on $[0, \infty)$,*

(iii) *$EZ_2 \leqq EZ_1 < \infty$,*

then

$$Z_1 \ll_c Z_2 \Rightarrow X/Z_1 \ll_0 X/Z_2.$$

Proof. The proof is an immediate consequence of Example 1 and Theorem 5(ii).

In closing this section on the properties of $\prec$ it may be of interest to note

Corollary 3. *The relation $\prec$ is antisymmetric; that is $Z_1 \prec Z_2$ and $Z_2 \prec Z_1 \Rightarrow Z_1$ and Z_2 have the same distribution.*

Proof. Antisymmetry follows from Corollary 1, since $Z_1 \prec Z_2$ and $Z_2 \prec Z_1$ implies that $g_{Z_1}(t) = g_{Z_2}(t)$ which, in turn, requires that Z_1 and Z_2 have the same probability distribution.

5. Generalized Hájek bounds. In this section we obtain two theorems which are generalizations of Hájek's Theorem 3. We first examine the set of random variables $D_Y = \{Z:Z \prec Y, Z \geqq 0\}$; that is, the set of all denominators which are uniformly better than Y.

Lemma 1. *The set D_Y is convex.*

Proof. We must show that $Z_1 \varepsilon D_Y$ and $Z_2 \varepsilon D_Y$ implies that $\alpha Z_1 + (1-\alpha)Z_2 \varepsilon D_Y$ for any $0 < \alpha < 1$. If both Z_1 and Z_2 are in D_Y, then by Theorem 4(ii) we have $Es(Z_i) \geqq Es(Y)$ for $i = 1, 2$ and all s in $\mathcal{S}$. Thus,

$$Es(\alpha Z_1 + (1-\alpha)Z_2) \geqq \alpha Es(Z_1) + (1-\alpha)Es(Z_2) \geqq Es(Y),$$

for all s in $\mathcal{S}$. The proof is complete.

Lemma 2. *Let $\mathcal{Z}_c = \{Y:EY \leqq c, Y \geqq 0\}$, and $Z' = c$ a.s., then $Z' \prec Y$ for all Y in $\mathcal{Z}_c$, and there is no other denominator Z with this property for which $Z' \prec Z$.*

Proof. This lemma actually states that Z' is a greatest lower bound for the set $\mathcal{Z}_c$. For any Y in $\mathcal{Z}_c$, we have $EY \leqq EZ'$, and clearly $Z' \ll_c Y$. Hence, by Theorem 5(ii), $Z' \prec Y$ for all Y in $\mathcal{Z}_c$. Since Z' itself is in $\mathcal{Z}_c$, any other variable Z with $Z \prec Y$ for all Y in $\mathcal{Z}_c$ must necessarily have $Z \prec Z'$.

We now use these lemmas to prove

Lemma 3. *Let $Z_1, Z_2, \cdots, Z_n$ be any n non-negative random variables with $Z_j \prec Z^*$ for $1 \leqq j \leqq n$. Then*

$$Z' \prec \textstyle\sum_{i=1}^n \lambda_i Z_i \prec Z^*, \quad \lambda_i \geqq 0, \quad \sum_{i=1}^n \lambda_i = 1, \tag{4}$$

where $Z' = \max_{1 \leqq i \leqq n} \{EZ_i\}$ a.s.

Proof. First note that $E(\sum_{i=1}^n \lambda_i Z_i) = \sum_{i=1}^n \lambda_i(EZ_i) \leqq EZ'$. From Lemma 2 we have $Z' \prec \sum_{i=1}^n \lambda_i Z_i$. Now consider the set $D_{Z^*} = \{Z:Z \prec Z^*\}$. By assumption, $Z_i \varepsilon D_{Z^*}$ for $1 \leqq i \leqq n$. By Lemma 1 D_{Z^*} is convex, and so all convex combinations of points of D_{Z^*} are members of D_{Z^*}. Thus, for $\lambda_i \geqq 0$ and $\sum_{i=1}^n \lambda_i = 1$, we have $\sum_{i=1}^n \lambda_i Z_i \prec Z^*$. The proof is complete.

In the case where all of the Z_i of Lemma 3 are independent and identically distributed we can sharpen the bounds given in (4). In fact, we can show

LEMMA 4. *Let $Z_1, Z_2, \cdots, Z_n$ be any n independent, identically distributed, non-negative random variables, then it follows that*

$$\sum_{i=1}^{n} (Z_i/n) \prec \sum_{i=1}^{n} \lambda_i Z_i \prec \sum_{i=1}^{\nu} (Z_i/\nu) \tag{5}$$

where the λ_i are as in Lemma 3 and ν is any integer not exceeding $\min_{1 \leq i \leq n} \{1/\lambda_i\}$.

PROOF. Given any $Z_{\lambda_0} = \sum_{i=1}^{n} \lambda_i^0 Z_i$, form the set $D_{Z_{\lambda_0}} = \{Z : Z \prec Z_{\lambda_0}\}$. Z_{λ_0} is determined by the weights $(\lambda_1^0, \lambda_2^0, \cdots, \lambda_n^0)$. Let $(\alpha_1^j, \alpha_2^j, \cdots, \alpha_n^j)$, $1 \leq j \leq n!$, be any of the $n!$ permutations of the λ_i^0. Because the Z_i are indentically distributed and independent, each of the $n!$ $Z_{\alpha_j} = \sum_{i=1}^{n} \alpha_i^j Z_i$ have the same distribution as Z_{λ_0}. But then we have

$$Z_{\alpha_j} \prec Z_{\lambda_0} \quad \text{for} \quad 1 \leq j \leq n!.$$

This tells us that all the Z_{α_j} are in $D_{Z_{\lambda_0}}$, and by Lemma 1 we have

$$\sum_{j=1}^{n!} \beta_j Z_{\alpha_j} \prec Z_{\lambda_0} \tag{6}$$

for all $\beta_j \geq 0$, $\sum_{j=1}^{n!} \beta_j = 1$. In particular, we have $\sum_{j=1}^{n!} (Z_{\alpha_j}/n!) = \sum_{i=1}^{n} (Z_i/n) \prec Z_{\lambda_0}$. This gives us the desired lower bound.

Assume that ν satisfies the specified condition. We can consider that $\sum_{i=1}^{\nu} (Z_i/\nu)$ has the form $\sum_{i=1}^{n} \lambda_i^0 Z_i$ where $(\lambda_1^0, \lambda_2^0, \cdots, \lambda_n^0) = (1/\nu, 1/\nu, \cdots, 1/\nu, 0, 0, \cdots, 0)$. Again all permutations of the λ_i^0 yield $Z_{\alpha_j} \in D_{Z_{\lambda_0}}$ and so once again (6) holds. A lemma, due to Hájek (1957), states that whenever the λ_i are such that $\nu \leq \min\{1/\lambda_i\}$, then there exists a $(\beta_1, \beta_2, \cdots, \beta_{n!})$ such that $\sum_{j=1}^{n!} \beta_j Z_{\alpha_j} = \sum_{i=1}^{n} \lambda_i Z_i$. Combining this with (6) gives us $\sum_{i=1}^{n} \lambda_i Z_i \prec \sum_{i=1}^{\nu} (Z_i/\nu)$ and this completes the proof.

In the case where nothing is known about the λ_i, the upper limit of the present lemma is the same as that given in Lemma 3. This comes about since, under these conditions $\min\{1/\lambda_i\} = 1$.

Lemmas 3 and 4 lead us to the following two generalizations of Theorem 3 (Hájek).

THEOREM 6. *Let (X, h) satisfy Condition (A), and let $\mathcal{Z} = \{Z : Z = \sum_{i=1}^{n} \lambda_i Z_i, \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1$, where the Z_i are any n non-negative random variables, independent of X, with $Z_j \prec Z_1$ for $2 \leq j \leq n\}$. Then*

$$X/h(\gamma) \ll_0 X/h(Z) \ll_0 X/h(Z_1) \tag{7}$$

for all Z in $\mathcal{Z}$, provided $\gamma = \max_{1 \leq i \leq n} \{EZ_i\}$.

PROOF. From Lemma 3 we know $Z \prec Z_1$ for all $Z \in \mathcal{Z}$. By assumption (X, h) satisfies Condition (A), and X is independent of Z and Z_1. It follows then from Definition 2 that $X/h(Z) \ll_0 X/h(Z_1)$. We also know that $Z' \prec Z$ for any Z in $\mathcal{Z}$, again by Lemma 3. Clearly Z' is independent of X, since it is a degenerate random variable. Thus, we have $X/h(\gamma) \ll_0 X/h(Z)$ for all Z in $\mathcal{Z}$, and the proof is complete.

In exactly the same manner we can use Lemma 4 to obtain the main result in Lawton (1965). Namely, THEOREM 7. *Let* (X, h) *satisfy Condition* (A), *and let* $\mathcal{Z} = \{Z : Z = \sum_{i=1}^{n} \lambda_i Z_i$, $\lambda_i \geqq 0$, $\sum_{i=1}^{n} \lambda_i = 1$, *where the* Z_i *are n independent, identically distributed, nonnegative random variables, all independent of* $X\}$. *Then*

$$X/h(\textstyle\sum_{i=1}^{n} Z_i/n) \ll_0 X/h(Z) \ll_0 X/h(\sum_{i=1}^{\nu} Z_i/\nu)$$

for all Z *in* $\mathcal{Z}$, *provided* $\nu = \min_{1 \leqq i \leqq n} \{1/\lambda_i\}$.

Notice that if one takes $Z_i = \chi_i^2(1)$, then this last theorem gives statement (2), and if one also asks that X have a standard normal distribution and takes $h(z) = z^{\frac{1}{2}}$, then this theorem becomes Theorem 3 of Hájek.

6. Applications to the Behrens-Fisher problem. Theorem 7 of the last section was used by Lawton (1965) to obtain bounds for the α and β risk in the Behrens-Fisher problem. The theorem was also used in that article to obtain α-risk bounds for the modified two-sample test, the u-test, proposed by Lord (1947) when the two sample sizes are equal.

We can extend the u-test result to the case of unequal sample sizes by using Theorem 6 in place of Theorem 7. Let $X_1, X_2, \cdots, X_{n_1}$ and $Y_1, Y_2, \cdots, Y_{n_2}$ be two samples from $N(\xi, \sigma^2)$ and $N(\eta, \tau^2)$ respectively. We will accept the hypothesis $H : \xi = \eta$ whenever

$$(8) \qquad |\bar{Y} - \bar{X}|/(W_1^2/n_1 + W_2^2/n_2)^{\frac{1}{2}} < c,$$

where W_1 and W_2 are the ranges of the respective samples. Under the null hypothesis the statistic in (8) has the form $X/h(Z)$ where the numerator has the standard normal distribution, $h(z) = z^{\frac{1}{2}}$, and Z is distributed as

$$\lambda W_1^2(n_1) + (1 - \lambda)W_2^2(n_2)$$

where $\lambda = (\sigma^2/n_1)/(\sigma^2/n_1 + \tau^2/n_2)$ and $W_i(n_i)$ denotes a random variable having the distribution of the range of a sample of size n_i from a standard normal population.

Now $W_1(n_1)$ and $W_2(n_2)$ are both independent of the numerator, and $W^2(n)$ is stochastically larger than $W^2(m)$ for $m \leqq n$, and so we may apply Theorem 6 to obtain

$$(9) \qquad P\{|V|/W(\nu) < c\} \leqq P\{|\bar{Y} - \bar{X}|/(W_1^2/n_1 + W_2^2/n_2)^{\frac{1}{2}} < c\} \leqq P\{|V|/\gamma^{\frac{1}{2}} < c\}$$

where V has the standard normal distribution, $\gamma = \max\{EW^2(n_1), EW^2(n_2)\}$, and $\nu = \min\{n_1, n_2\}$. If one takes $n_1 = n_2 = 10$ and $c = 0.288$, then the bounds for the probability of Type I error are $0.0036 \leqq \alpha \leqq 0.02$. The lower bound given here by the statistic $|V|/\gamma^{\frac{1}{2}}$ is actually better than the bound given in Lawton (1965) by the statistic $|(2/n)^{\frac{1}{2}} \cdot V|/(W_1(n_1) + W_2(n_2))$.

In the present case, because the $W_i(n)$ are independent and stochastically increasing in n, we can further improve the upper bound given in (9).

Theorem 6 told us only that $Z' < \lambda W^2(n_1) + (1 - \lambda)W^2(n_2)$, where Z' was degenerate at γ. Take $m = \max\{n_1, n_2\}$, then $W^2(m)$ is stochastically greater than $W_i^2(n_i)$ for $i = 1, 2$. It follows then, since the $W_i^2(n_i)$ are independent, that $\lambda W_1^2(m) + (1 - \lambda)W_2^2(m)$ is stochastically larger than $\lambda W_1^2(n_1) + (1 - \lambda)W_2^2(n_2)$. The $W_i^2(m)$ are considered to be independent. Thus, we know

$$\lambda W_1^2(m) + (1 - \lambda)W_2^2(m) \prec \lambda W_1^2(n_1) + (1 - \lambda)W_2^2(n_2).$$

But the left hand quantity involves the average of two independent, identically distributed random variables, both independent of X. We can apply Theorem 7 to obtain

$$[(W_1^2(m) + W_2^2(m))/2] \prec \lambda W_1^2(n_1) + (1 - \lambda)W_2^2(n_2).$$

Combining this with (8) we have

$$(10) \qquad P\{|V|/W(\nu) < c\} \leqq P\{|\bar{Y} - \bar{X}|/[W_1^2/n_1 + W_2^2/n_2]^{\frac{1}{2}} < c\}$$
$$\leqq P\{(2/n)^{\frac{1}{2}}|V|/[W_1^2(m) + W_2^2(m)]^{\frac{1}{2}} < c\}$$

where $\nu = \min\{n_1, n_2\}$ and $m = \max\{n_1, n_2\}$. The distribution of the statistic in the upper bound is unfortunately not tabulated. If we take $n_1 = n_2$ then the above inequality is the same as that in Lawton (1965).

7. Confidence intervals. One could also use Theorems 6 and 7 to construct confidence intervals for linear combinations of means from k normal populations.

Let $X_{i1}, X_{i2}, \cdots, X_{in_i}$ be the ith sample of size n_i from $N(\mu_i, \sigma_i^2)$ where $i = 1, 2, \cdots, k$. Suppose one wishes to construct a confidence interval for $\theta = \sum_{i=1}^k \gamma_i\mu_i$ where the γ_i are known constants, but the σ_i^2 are unknown, and possibly unequal. Such confidence intervals may be obtained from the statistic

$$(11) \qquad [\textstyle\sum_{i=1}^k \gamma_i X_{i\cdot} - \theta]/[\sum_{i=1}^k (\gamma_i^2 s_i^2/n_i)]^{\frac{1}{2}}$$

where $X_{i\cdot} = \sum_{j=1}^{n_i} X_{ij}/n_i$, and $s_i^2 = \sum_{j=1}^{n_i} (X_{ij} - X_{i\cdot})^2/(n_i - 1)$.

On the division of numerator and denominator by $[\sum_{i=1}^k (\gamma_i^2\sigma_i^2/n_i)]^{\frac{1}{2}}$ the statistic in (11) has the form $X/h(Z)$ with X standard normal, and $h(z) = z^{\frac{1}{2}}$. Z is independent of X and has the form

$$(12) \qquad \textstyle\sum_{i=1}^k \lambda_i\chi_i^2(m_i)/m_i$$

where $\lambda_i = (\gamma_i^2\sigma_i^2/n_i)/[\sum_{i=1}^k (\gamma_i^2\sigma_i^2/n_i)]$ and $m_i = n_i - 1$. As in Section 6, this can be put in the form $\sum_{i=1}^m \alpha_i\chi_i^2(1)$, $\alpha_i \geqq 0$, $\sum_{i=1}^m \alpha_i = 1$, where $m = \sum_{i=1}^k n_i - k$. The $\chi_i^2(1)$ are all independent of the numerator, and so the conditions for Theorem 7 are satisfied. We are then able to obtain bounds for the probability that the statistic in (11) lies in an interval about zero. This allows the construction of confidence intervals for θ based on the means and standard deviations of the k samples. In much the same way, Theorem 6 would allow one to obtain confidence intervals for θ based on the statistic

$$[\textstyle\sum_{i=1}^k \gamma_i X_{i\cdot} - \theta]/[\sum_{i=1}^k (\gamma_i^2 W_i^2/n_i)]^{\frac{1}{2}}$$

where the sample range of the ith sample replaces s_i as the estimator of σ_i. If the sample sizes are equal, one could just as well use any of the usual estimators for σ_i; interquartile range, mean average deviation, etc.

An interesting application of this method is given by the following example. Three sets of measurements are made on the same object, and it is felt that the precision of the three sets of measurements may not be the same. That is, $X_{i1}, X_{i2}, \cdots, X_{in_i}$ are samples of size n_i from $N(\mu, \sigma_i^2)$ for $i = 1, 2, 3$. Let $n = n_1 + n_2 + n_3$. It is possible to obtain confidence intervals for μ based on the overall mean $\bar{X} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} (X_{ij}/n)$. This is done by choosing $\gamma_i = n_i/n$ and using the statistic in (11). In this case, $\theta = \mu$ and the confidence intervals may be obtained from the inequality

$$t_{n-3} \ll_0 (\bar{X} - \mu)/[\sum_{i=1}^{3} (n_i s_i^2/n^2)]^{\frac{1}{2}} \ll_0 t_\nu$$

where $\nu = \min\{n_1 - 1, n_2 - 1, n_3 - 1\}$ and t_m denotes a random variable having the central t-distribution with m degrees of freedom. If we had assumed that all the variances were equal, then the confidence intervals could have been obtained from a statistic having the t-distribution with $n - 1$ degrees of freedom.

We can also, of course, use the statistic in (11) for testing hypotheses about θ. The bounds we have obtained for the confidence coefficient then become bounds for the type I error. As in Lawton (1965), we could extend these bounds to the power curve for values of the non-centrality parameter $|\delta| \leqq 2$. We could then construct tests involving θ which would guarantee that type I error is at most α_0, while the power at alternative δ_0 is at least β_0.

8. Randomization theorem. In this section we shall show that $Z_1 \prec Z_2$ if and only if Z_2 can be obtained from Z_1 by a particular type of randomization. While this result has been stated in Theorem 4 of Section 4, its proof has been deferred to now because of its length and complexity. The present results come from a modification of some theorems given by Meyer (1963) on the characterization of the Choquet ordering of positive measures.

In this section let $\bar{R}^+$ again be the compact, positive half-line, and let $\mathcal{C}$ be the space of continuous functions on $\bar{R}^+$. Let $\mathcal{R}$, $\mathfrak{M}^+$, and $\mathfrak{M}$ denote the classes of Radon measures, sub-stochastic measures, and probability measures respectively. A sub-stochastic measure is a positive measure with total mass at most one. Define Δ_x to be the distribution function of the measure which assigns mass 1 to the point x in $\bar{R}^+$. For any $F_Z \in \mathcal{R}$ define

$$EZ = \int_{\bar{R}^+} y\, dF_Z(y)$$

provided the integral exists. Finally, let $\mathcal{S}$ be the space of all non-negative, non-decreasing, concave functions belonging to $\mathcal{C}$. From Theorem 4 we know

$$Z_1 \prec Z_2 \Leftrightarrow Es(Z_1) \geqq Es(Z_2)$$

for all s in $\mathcal{S}$. Here Z_i denotes a random variable.

DEFINITION 4. For any $f \in \mathcal{C}$ define $\hat{f}$ by

$$\hat{f}(x) = \inf\{s(x) : s \in \mathcal{S},\ s \geqq f\}$$

for each x in $\bar{R}^+$.

$\hat{f} \geqq f$ and is concave, and continuous.

THEOREM 8. *For all x in $\bar{R}^+$ and each f in $\mathcal{C}$*

$$\hat{f}(x) = \sup\{Ef(Z) : F_Z \varepsilon \mathfrak{M}^+, EZ \leqq x\}.$$

PROOF. If $EZ \leqq x$, then $Es(Z) \leqq s(EZ) \leqq s(x)$ for all s in $\mathcal{S}$. The first inequality follows from Jensen's inequality, and the second from the fact that s is a non-decreasing function. But then $\hat{f}(x) = \inf\{s(x) : s \varepsilon \mathcal{S}, s \geqq f\} \geqq \sup\{Es(Z) : s \varepsilon \mathcal{S}, s \geqq f\} \geqq Ef(Z)$ for any $F_Z \varepsilon \mathfrak{M}^+$ with $EZ \leqq x$. We have then

$$\hat{f}(x) \geqq \sup\{Ef(Z) : F_Z \varepsilon \mathfrak{M}^+, EZ \leqq x\}.$$

In order to establish the reverse inequality, we note that the map defined by $L_x(f) = \hat{f}(x)$ is sub-linear. That is,

$$(13) \qquad \begin{aligned} L_x(af) &= aL_x(f) && \text{for } a > 0, \\ L_x(f + g) &\leqq L_x(f) + L_x(g) && \text{for } f, g \varepsilon \mathcal{C}. \end{aligned}$$

The first property in (13) follows immediately from the definition of $\hat{f}$. Also, we see that

$$\begin{aligned} L_x(f + g) &= \inf\{s(x) : s \geqq f + g, s \varepsilon \mathcal{S}\} = \inf \mathcal{S}_1, \\ L_x(f) + L_x(g) &= \inf\{s_1(x) + s_2(x) : s_1 \geqq f, s_2 \geqq g, s_i \varepsilon \mathcal{S}\} \\ &= \inf \mathcal{S}_2. \end{aligned}$$

But since $s_1 + s_2 \varepsilon \mathcal{S}$ we have $\mathcal{S}_2 \subset \mathcal{S}_1$, and so

$$L_x(f + g) = \inf \mathcal{S}_1 \leqq \inf \mathcal{S}_2 = L_x(f) + L_x(g).$$

Using this we have

$$(14) \qquad L_x(f) = \sup\{Ef(Z) : F_Z \varepsilon \mathcal{R}, Ef(Z) \leqq L_x(f)\};$$

that is, L_x is the upper envelope of linear functionals on $\mathcal{C}$ less than L_x. Let b be the supremum in (14). If $b < L_x(f)$ then $b + \epsilon = L_x(f)$ for some $\epsilon > 0$. If (14) were not true then by use of the Hahn-Banach theorem one could construct a linear functional on $\mathcal{C}$ whose value at f is $b + \epsilon/2 < L_x(f)$. This contradicts the statement that b was the supremum of such linear forms.

Let $Ef(Z)$ be any one of these linear functionals which lie below $L_x(f)$. If $f \leqq 0$, then $\hat{f} \leqq 0$, and this implies $Ef(Z) \leqq \hat{f}(x) \leqq 0$ since $s(x) = 0$ is in $\mathcal{S}$. This last statement implies that F_Z must define a positive measure. We also have $f(x) = 1$ in $\mathcal{S}$ and this yields $E(1) \leqq \hat{f}(x) = 1$, and $E(1)$ is the total mass associated with the measure F_Z. Thus, we have shown that $F_Z \varepsilon \mathfrak{M}^+$. Finally, note that $f(x) = x$ is in $\mathcal{S}$ and this yields $EZ \leqq \hat{f}(x) = x$. We have then shown that

$$\begin{aligned} \hat{f}(x) &= \sup\{Ef(Z) : F_Z \varepsilon \mathcal{R}, Ef(Z) \leqq \hat{f}(x)\} \\ &\leqq \sup\{Ef(Z) : F_Z \varepsilon \mathfrak{M}^+, EZ \leqq x\}. \end{aligned}$$

This inequality completes the proof.

Let U be the compact set in $\mathfrak{M} \times \mathfrak{M}$ (having the topology induced by $\mathcal{C}$)

defined by

$$U = \{(\Delta_x, F_Z): x \varepsilon \bar{R}^+, F_Z \varepsilon \mathfrak{M}, EZ \leqq x\}.$$

It is clear that $U \subset V$, the closed convex set defined by

$$V = \{(F_{Z_1}, F_{Z_2}): (F_{Z_1}, F_{Z_2}) \varepsilon \mathfrak{M} \times \mathfrak{M}, Z_1 \prec Z_2\}.$$

LEMMA 5. *V is the closed convex envelope of U.*

PROOF. Let U_c denote the closed convex envelope of U. We know $U_c \subset V$, and we need to show that there are no points of V outside U_c. By virtue of the Hahn-Banach theorem, it will be enough to show the following:

If $(F_{Z_1}, F_{Z_2}) \varepsilon V$ and L is any continuous linear functional on $\mathfrak{M} \times \mathfrak{M}$ with $U \subset \{L \geqq 0\}$, then $V \subset \{L \geqq 0\}$. That is, all linear half-spaces containing U also contain V.

All continuous linear functionals on $\mathfrak{M} \times \mathfrak{M}$ have the form

$$L(H, G) = \int u\, dH - \int v\, dG, \qquad (u, v) \varepsilon \mathfrak{C} \times \mathfrak{C}.$$

Thus, we seek to show that

$$L(\Delta_x, F_Z) = u(x) - Ev(Z) \geqq 0 \quad \text{for every} \quad (\Delta_x, F_Z) \varepsilon U$$

implies that

$$Eu(Z_1) - Ev(Z_2) \geqq 0, \qquad \text{for all} \quad (F_{Z_1}, F_{Z_2}) \varepsilon V.$$

But $L(\Delta_x, F_Z) \geqq 0$ for all (Δ_x, F_Z) in U implies that

$$u(x) \geqq Ev(Z) \qquad \text{for all} \quad E_Z \varepsilon \mathfrak{M}^+, EZ \leqq x,$$

which in turn yields

$$u(x) > \hat{v}(x) \qquad \text{for all} \quad x \quad \text{in} \quad \bar{R}^+.$$

But then $Eu(Z_1) \geqq E\hat{v}(Z_1) \geqq E\hat{v}(Z_2) \geqq Ev(Z_2)$ as was to be shown. The proof is complete.

THEOREM 9. *If $Z_1 \prec Z_2$, then there exists a probability measure θ on U whose center of gravity (in $\mathfrak{M} \times \mathfrak{M}$) is equal to (F_{Z_1}, F_{Z_2}).*

PROOF. The result follows immediately from the last lemma, and the fact that U_c is a closed, convex, compact set in $\mathfrak{M} \times \mathfrak{M}$. U_c is then, by the Krein-Milman theorem, the set of centers of gravity of probability measures on U. The proof is complete.

We are now ready to prove the main theorem of this section.

THEOREM 10. *$Z_1 \prec Z_2$ if and only if there exists a randomized, mean decreasing map T such that $Z_2 = T(Z_1)$.*

PROOF. That this condition is sufficient follows easily. Let T be an rmd map and $F_{Z_1} \varepsilon \mathfrak{M}$, then $T_{z_1}(z_2)$ can be interpreted as the conditional distribution of Z_2 given Z_1. Since $E(Z_2 \mid z_1) = \int y\, dT_{z_1}(y) \leqq z_1$, we have

$$Es(Z_2) = E[Es(Z_2 \mid Z_1)] \leqq E[s(E(Z_2 \mid Z_1))] \leqq Es(Z_1)$$

for all s in $\mathcal{S}$. By Theorem 4 we have $Z_1 \prec Z_2$.

Conversely, let $(F_{Z_1}, F_{Z_2}) \varepsilon \mathfrak{M} \times \mathfrak{M}$ and such that $Z_1 < Z_2$. By Theorem 9 there exists a probability measure θ on U such that

$$(15) \qquad \int\int_U (\Delta_x, F)\theta(d\Delta, dF) = (F_{Z_1}, F_{Z_2}).$$

We can define a new measure θ' on $U' = \{(x, F): x \varepsilon \bar{R}^+, F \varepsilon \mathfrak{M}\}$ by means of the map $(\Delta_x, F) \rightarrow (x, F)$. We may rewrite (15) as

$$\int\int_{U'} (\Delta_x, F)\theta'(dx, dF) = (F_{Z_1}, F_{Z_2}).$$

Using the notion of regular conditional probabilities we can write

$$(16) \qquad \int\int_{U'} (\Delta_x, F)\theta'(dF \mid x)\theta'(dx) = (F_{Z_1}, F_{Z_2})$$

where $\theta'(dF \mid x)$ represents the conditional distribution of F given x, and $\theta'(dx)$ the marginal distribution of x. These measures are determined by the joint distribution θ'.

From (16) we have

$$\int_{\bar{R}^+} \Delta_x(y)\theta'(dx) = F_{Z_1}(y),$$

and this is equivalent to saying $F_{Z_1}(y)$ is the distribution function corresponding to the measure $\theta'(dx)$. Also we have

$$\int_{\bar{R}^+} [\int F(y)\theta'(dF \mid x)] \, dF_{Z_1}(x) = F_{Z_2}(y).$$

Let $T_x(y)$ be the quantity in square brackets. Being an average of distribution functions of probability measures on $\bar{R}^+$ for each fixed x, it is itself the distribution function of a probability measure on $\bar{R}^+$. In addition, $\int y \, dF(y) \leqq x$ for every F in the average, implies $\int y \, dT_x(y) \leqq x$. Finally, one sees that, for each fixed y, $T_x(y)$ is a measurable function of x. It follows that $T_x(y)$ is the rmd map which was sought.

Acknowledgment. I wish to thank Dr. Erich Lehmann for suggesting the possibility of extending Hájek's work, and Dr. Lucien LeCam for his suggestions which led to Section 8 of this paper. My thanks also for their many helpful suggestions during the preparation of this paper.

REFERENCES

[1] Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Statist.* **19** 76–81.

[2] Hájek, J. (1962). Inequalities for the generalized Student's distribution. *Sel. Transl. Math. Statist. Prob.* **2** 63–74.

[3] Lawton, W. H. (1965). Some inequalities for central and non-central distributions. *Ann. Math. Statist.* **36** 1521–1525.

[4] Lord, E. (1947). The use of the range in place of the standard deviation in the t-test. *Biometrika* **34** 41–67.

[5] Lord, E. (1950). Power of the modified t-test (u-test) based on range. *Biometrika* **37** 64–77.

[6] Meyer, P. A. (1963). *Progrès Récents dans la Théorie des Cônes à Base Compacte.* Seminaire Brelot-Choquet-Deny Théorie du Potentiel. 1963/64.