

GENERALIZED CONVEX INEQUALITIES

SAMUEL KARLIN AND ALBERT NOVIKOFF

This paper concerns certain cones of functions φ and their duals. The functions of these cones can be described globally by determinantal inequalities with the aid of certain auxiliary functions ψ_1, \dots, ψ_k , or locally, if φ is sufficiently smooth, by differential inequalities. In the latter case smoothness hypotheses are imposed on ψ_1, \dots, ψ_k in order to define the relevant differential operator. The cones considered can be regarded as generalizations of the classical cone of convex functions. Various special cases occur in moment theory, problems of interpolation, differential equations, probability theory and elsewhere.

The elements of the corresponding dual cones are measures $d\mu$ (not necessarily nonnegative) which are orthogonal to ψ_1, \dots, ψ_k , and in addition satisfy an inequality described with the aid of an integral operator (see (20) below). A sufficient condition for membership in the dual is that $d\mu$ have the minimal oscillation compatible with being orthogonal to ψ_1, \dots, ψ_k , a condition depending on k but not otherwise on the choice of ψ_1, \dots, ψ_k (see Lemma B).

In the first section we consider the classical case of the cone of convex functions (and their discrete analogues, convex sequences). This case is presented in some detail, despite its familiarity and elementary character, in order to motivate the remainder. Much of this section was anticipated in [3] and elsewhere in the subsequent literature, but our viewpoint is somewhat different.

We introduce the general cones $C(\psi_1, \dots, \psi_k)$ in § 3 where for ease of exposition we restrict ourselves to smooth ψ_1, \dots, ψ_k . In this case the associated differential operator is easy to describe. We also impose strictness in certain constraining inequalities satisfied by ψ_1, \dots, ψ_k . A characterization of the dual cone and the corresponding sufficient condition for membership in it are then given (Theorem 1 and Lemma B). As applications we list several known inequalities of previously unrelated character, and obtain some new ones. In § 5 we weaken some of the strictness conditions referred to above by introducing a useful procedure for approximating a cone with weakened hypotheses by one with strictness. In § 4 we exploit the translation invariant character of the cone whose associated differential operator is itself translation invariant (i.e. linear with constant coefficients).

Evidently we can weaken the hypothesis of smoothness concerning ψ_1, \dots, ψ_k and we can extend the considerations of § 4 to more gener-

Received August 29, 1962, and in revised form May 1, 1963. Supported in part by National Science Foundation Grant 16319.

al cones by introducing the notion of translation invariant with respect to more general differential operators. A full treatment of these matters and others is planned for a future publication. For example, we will describe fully the structure of the extreme rays of $C(\psi_1, \dots, \psi_k)$ and its dual. We also extend our considerations to cones which are the intersections $C(\psi_1) \cap C(\psi_1, \psi_2) \cap \dots \cap C(\psi_1, \dots, \psi_k)$ of those studied here and obtain corresponding inequalities (e.g., for functions which are positive, increasing and convex).

1. **Ordinary convexity, continuous case.** We consider the class K of convex functions φ defined on a finite interval $[a, b]$ and ask what measures $d\mu$ (in this context measures are not necessarily non-negative) on $[a, b]$ satisfy

$$(1) \quad \int_a^b \varphi d\mu \geq 0,$$

for all φ in K . We call the set of such measures K^* and note that K and K^* are convex cones. The restriction that $[a, b]$ be finite is inessential, and with proper interpretation of various possibly divergent expressions, the results that follow remain true even if $[a, b]$ is infinite. We refer the reader to [5] for a full treatment of the corresponding modifications. (A function φ is said to be convex in $[a, b]$ if

$$(2) \quad \begin{vmatrix} 1 & x_1 & \varphi(x_1) \\ 1 & x_2 & \varphi(x_2) \\ 1 & x_3 & \varphi(x_3) \end{vmatrix} \geq 0 \quad \text{for } a \leq x_1 < x_2 < x_3 \leq b.)$$

It is easily seen that the C^2 elements of K are weakly dense in K , so that it is enough to investigate (1) for $\varphi \in C^2 \cap K$. If $d\mu$ is in K^* then clearly

$$(i) \quad \int_a^b d\mu = 0$$

and

$$(ii) \quad \int_a^b x d\mu = 0$$

since $\varphi_1(x) = 1$ and $\varphi_2(x) = x$ satisfy $\pm\varphi_1 \in K$, $\pm\varphi_2 \in K$. It follows that in considering $\int_a^b \varphi d\mu$ we may assume $\varphi(a) = 0$, $\varphi'(a) = 0$, and $\varphi \in C^2 \cap K$.

Let $\mu_1(x) = \int_a^x d\mu$ and $\mu_2(x) = \int_a^x \mu_1(t) dt$. Then (i) implies $\mu_1(b) = 0$,

and (i) and (ii) together imply $\mu_2(b) = 0$ so that integrating by parts twice we have the identity:

$$(A) \quad \int_a^b \varphi d\mu = \int_a^b \varphi'' \mu_2(x) dx$$

A consequence of (A) is that the condition

$$(iii) \quad \mu_2(x) \geq 0 \quad \text{in } [a, b]$$

is necessary in order that $d\mu \in K^*$. However, conditions (i), (ii), and (iii) together are in fact sufficient as well: for $\varphi \in C^2$ and μ satisfying (i) and (ii) the identity (A) holds, so that (i), (ii), and (iii) show that $\int_a^b \varphi d\mu \geq 0$ for the C^2 member of K , and thus for arbitrary φ in K .

We single out the special case of absolutely continuous measures $d\mu = f(x)dx$, for which the corresponding necessary and sufficient conditions are explicitly:

$$(i)^* \quad \int_a^b f(x) dx = 0$$

$$(ii)^* \quad \int_a^b xf(x) dx = 0$$

$$(iii)^* \quad \int_a^x \int_a^\xi f(t) dt d\xi \geq 0 \quad a \leq x \leq b .$$

(We remark again that equality holds in (iii)* for $x = b$ as a consequence of (i)* and (ii)*.)

We claim no novelty for this result, which is implicit in [3] and undoubtedly can be found elsewhere.

We now state two simple lemmas which are well-known, and whose proof is in any case contained in the results of § 3.

LEMMA a. *If $d\mu$ is non-null and satisfies (i) and (ii), then $[a, b]$ must have at least two intervals on which $d\mu(x)$ (or its negative) is nonnegative separated by one on which $d\mu$ is nonpositive and in each of which it is nonnull. We summarize this state of affairs by saying that $d\mu$ has at least 2 strict changes of sign. (An atom of nonzero measure is also considered an interval, cf. Example 3.)*

LEMMA b. *If $d\mu$ satisfies (i) and (ii), then a sufficient condition that $d\mu \in K^*$ is that $d\mu(x)$ have exactly two strict changes of sign and be "positive near a " (i.e., be nonnegative and nonnull in some interval $[a, x_1]$, $a \leq x_1 < b$).*

The point of Lemma b is that its sign-change property (which

we may call the "minimal sign-change" property in virtue of Lemma a) is frequently easier to verify than condition (iii). It is useful to observe that the sign change property is invariant under monotone changes of variable. When $d\mu(x) = f(x)dx$ then of course the meaning of sign-change for $d\mu$ reduces to the familiar one for the function $f(x)$.

Applications.

(1) Favard's Inequality [1]. Let $f(x)$ be a nonnegative continuous concave function in $[a, b]$, not identically 0, and let $\psi(y)$ be convex in $[0, 2\bar{f}]$, where $\bar{f} = (b - a)^{-1} \int_a^b f(x)dx$. Then:

$$(F) \quad \frac{1}{2\bar{f}} \int_0^{2\bar{f}} \psi(y)dy \geq \frac{1}{b - a} \int_a^b \psi[f(x)]dx .$$

It is sufficient to establish (F) under the assumption that f is C^1 and strictly concave; the general case then follows by a limit procedure. We introduce the measure $d\mu_f$ defined for all continuous functions θ by

$$\frac{1}{b - a} \int_a^b \theta(f(x))dx = \int_0^\infty \theta(y)d\mu_f .$$

In the case under consideration, $d\mu_f(y) = (1/(b - a))m(y)dy$ where $m(y)$ is given by the increasing function $\sum (|f'(x)|)^{-1}$ for y in $[f_{\min}, f_{\max}]$ (and zero elsewhere), the sum extending over the at most two values of x satisfying $f(x) = y$.

We now consider the difference:

$$\frac{1}{2\bar{f}} \int_0^{2\bar{f}} \psi(y)dy - \frac{1}{b - a} \int_a^b \psi[f(x)]dx = \int_0^\infty \psi(y)[d\nu(y) - d\mu_f(y)]$$

where

$$\nu(y) = \begin{cases} \frac{y}{2\bar{f}} & 0 \leq y \leq 2\bar{f} \\ 1 & 2\bar{f} \leq y . \end{cases}$$

It is immediate from the choice of $\nu(y)$ that this difference vanishes for $\psi_1(y) \equiv 1$, $\psi_2(y) \equiv y$, so that (i) and (ii) with respect to the interval $[0, 2\bar{f}]$ are satisfied by $d\nu(y) - d\mu_f(y)$. By Lemma a, $d\nu(y) - d\mu_f(y)$ must have at least 2 sign changes (implying $f_{\max} \leq 2\bar{f}$, a result also stated by Favard). Since $m(y)$ is increasing it is easy to see that $d\nu(y) - d\mu_f(y)$ has exactly two sign changes, in the order $+, -, +$. It follows by Lemma b that (F) holds.

A similar argument proves

$$(F') \quad \frac{1}{2\bar{f} - 2c} \int_c^{2\bar{f}-c} \psi(y)dy \geq \frac{1}{b-a} \int_a^b \psi(f(x))dx$$

for any $0 < c \leq f_{\min}$ where $\psi(y)$ is convex on the interval $[c, 2\bar{f} - c]$.

With the guide presented by (F') we can assert a version of (F') applicable in the case where $f(x)$ is continuous nonnegative and convex on $[a, b]$. A similar analysis as above shows that when f is convex then $m(y)$ for $d\mu_f(y) = 1/(b-a)m(y)dy$ is decreasing on $[f_{\min}, f_{\max}]$. Let $d > f_{\min}$. The difference $d\nu(y) - d\mu_f(y)$ ($d\nu(y) = (2\bar{f} - 2d)^{-1}$ on $[d, 2\bar{f} - d]$; $= 0$ elsewhere) exhibits the sign change arrangement $-, +, -$ as y traverses the positive axis. This leads with the aid of Lemma b to the inequality

$$(F'') \quad \frac{1}{2\bar{f} - 2d} \int_d^{2\bar{f}-d} \psi(y)dy \leq \frac{1}{b-a} \int_a^b \psi(f(x))dx$$

when $\psi(y)$ is convex on $[d, 2\bar{f} - d]$. Note that this inequality is in the opposite direction to that of (F). We also emphasize that c and d can be specified arbitrarily provided only that $d > f_{\min} > c$. Favard's case (F) is special corresponding to $c = 0$. The better formulation appears to be our extensions (F') and (F'').

(2) If $F(x)$ is a distribution of a nonnegative random variable (i.e., dF is nonnegative with support $[0, \infty]$ and total measure 1) with finite mean $\mu = \int_0^\infty x dF = \int_0^\infty (1 - F(x))dx$. Suppose further that $dF = e^{-\psi(x)}dx$ where ψ is convex. Then:

$$(4) \quad \int_0^\infty \varphi(x) dF(x) \leq \int_0^\infty \varphi(x) e^{-x/\mu} \frac{dx}{\mu}$$

for convex φ in $[0, \infty]$ (both sides may be infinite). This is again a new result (cf. Example 5 of § 2). If $d\mu = dG - dF$ where $dG = e^{-x/\mu}(dx/\mu)$ in $[0, \infty]$ then (i) and (ii) are clearly satisfied. The difference $e^{-\psi(x)} - (1/\mu)e^{-x/\mu}$ has the signature of $x/\mu + \log \mu - \psi(x)$ so that Lemma b applies to give the desired result. It is entirely elementary that if $dF = f(x)dx$ where:

$$\begin{vmatrix} f(x_1 - y_1) & f(x_1 - y_2) \\ f(x_2 - y_1) & f(x_2 - y_2) \end{vmatrix} \geq 0 \quad \text{for } 0 \leq x_1 < x_2, 0 \leq y_1 < y_2,$$

then $f(x) = e^{-\psi(x)}$ for convex ψ and conversely. Such densities are called Pólya frequency of order 2 (PF_2) on $[0, \infty]$ and are of independent interest in statistical applications [8]. It should be remarked, that this example applies without modification to comparison of densities of the form $dF = h[\psi(x)]dx$, $dG = h(x)dx$ whenever $\int_0^\infty (1 - F)dx = \int_0^\infty (1 - G)dx$, ψ is convex and h is any strictly decreasing function.

(3) Let F be a distribution function as in Example 2, but not restricted as to the support of its corresponding measure dF . Then for all convex φ

$$(5) \quad \varphi(\mu) \leq \int_{-\infty}^{+\infty} \varphi(x) dF(x)$$

(Jensen's inequality, cf. [1]).

Indeed, if

$$G(x) = \begin{cases} 0 & x < \mu \\ 1 & x \geq \mu \end{cases}$$

then it is trivial to verify Lemma b (or indeed (i)–(iii)) for $d\mu = dF(x) - dG(x)$. Since:

$$\varphi(\mu) = \int_{-\infty}^{+\infty} \varphi(x) dG(x)$$

this gives us the desired result.

2. Ordinary, convex, discrete case. Consider at first the cone of finite sequences $\{a_0, \dots, a_n\}$ such that

$$(6) \quad \sum_0^n a_k \varphi(k) \geq 0$$

for every convex sequence $\varphi(k)$, $k = 0, 1, \dots, n$.

It is easy to see that this discrete analogue of the previous section is subsumed by the consideration there, and corresponds to the subclass of purely atomic measures $d\mu$ which have their support on the integers $k = 0, 1, \dots, n$, with associated masses a_0, \dots, a_n . The conditions

$$(i)' \quad \sum_0^n a_k = 0$$

$$(ii)' \quad \sum_0^n k a_k = 0$$

$$(iii)' \quad \sum_{j=0}^k \sum_{i=0}^j a_i \geq 0, \quad k = 0, 1, \dots, n$$

are necessary and sufficient for (6) to hold for all convex φ . We remark that there is no need to restrict ourselves to finite series, but the interpretation of $\sum_0^\infty a_k \varphi(k) \geq 0$ for convex $\varphi(k)$ is that this have the usual meaning if the series is convergent, and that the left-hand side, if divergent, diverges to $+\infty$.

The following lemmas are easily verified:

LEMMA a'. If (i)' and (ii)' hold, then $\{a_0, \dots, a_n\}$ has at least two changes of sign.

LEMMA b'. If (i)' and (ii)' hold, and the signatures of a_0, \dots, a_n are $+, -, +$ in the sense that:

$$(7) \quad \begin{aligned} a_k &\geq 0 && \text{for } 0 \leq k \leq k_1 \\ a_k &\leq 0 && \text{for } k_1 + 1 \leq k \leq k_2 \\ a_k &\geq 0 && \text{for } k_2 + 1 \leq k \leq n \end{aligned}$$

with strict inequality holding at least once in each of the three indicated regions, then (6) holds for all convex $\varphi(k)$, $k = 0, 1, \dots, n$.

Applications.

(1) Let $0 \leq p_i \leq 1$, $i = 1, \dots, n$, $p_i + q_i = 1$ and define $\{b_k\}$ by

$$\sum_0^n b_k x^k = \prod_{i=1}^n (q_i + p_i x).$$

Further, let:

$$d_k = \binom{n}{k} p^k (1 - p)^{n-k}$$

i.e.,

$$\sum_0^n d_k x^k = (q + px)^n$$

where $p = (1/n)(p_1 + \dots + p_n)$. Then

$$(8) \quad \sum_0^n b_k \varphi(k) \leq \sum_{k=0}^n d_k \varphi(k)$$

for all convex $\varphi(k)$. This result is due to Hoeffding [4] and can be interpreted as an inequality of the form:

$$(9) \quad E_x(\varphi) \leq E_Y(\varphi)$$

for convex φ . ($E_x(\varphi)$ denotes the expectation of the function with respect to the random variable X .) Here X and Y are random variables denoting the number of successes in n independent binomial trials of probability p_i ($i = 1, \dots, n$) and n independent binomial trials each of probability p respectively.

It was this result which motivated the present investigation. For the proof we need only show that $a_k = d_k - b_k$ satisfies (i)'-(iii)'. However, we establish rather more in the next example.

(2) Let $0 \leq p_i \leq 1$, $i = 1, \dots, n$ be as in Example 1. Let π_{ij} be

a doubly stochastic matrix $0 \leq i, j \leq n$ (i.e., $\pi_{ij} \geq 0$, $\sum_{i=1}^n \pi_{ij} = \sum_{j=1}^n \pi_{ij} = 1$) and let $p'_i = \sum_{j=1}^n \pi_{ij} p_j$, $i = 1, \dots, n$. Let b_k be defined as in Example 1 and b'_k be defined similarly by $\sum_0^n b'_k x^k = \prod_{i=1}^n (q'_i + p'_i x)$. If Y' is the random-variable consisting of the number of successes in n independent binomial trials of probability p'_i ($i = 1, \dots, n$), then $E_{Y'}(\varphi) \leq E_Y(\varphi)$ for convex φ , or equivalently

$$(10) \quad \sum_{k=0}^n \varphi(k)(b'_k - b_k) \geq 0$$

for convex φ . Example 1 is obtained by an obvious specialization of the matrix π_{ij} . We assert that it suffices to establish this in the case:

$$\begin{aligned} p'_1 &= tp_1 + (1 - t)p_2 \\ p'_2 &= (1 - t)p_1 + tp_2 & 0 < t < 1 \\ p'_i &= p_i & i = 3, 4, \dots, n. \end{aligned}$$

If the matrix of this transformation is replaced by one in which the 1st two rows and columns are interchanged with arbitrary rows and columns, the result still holds, since only the order of the $\{p_i\}$ and $\{p'_i\}$ are changed. However, as shown in [2 Chapter 1], an arbitrary doubly stochastic matrix is a finite product of just such matrices, with differing choices of t .

A simple computation now shows that:

$$b'_k - b_k = (p'_1 p'_2 - p_1 p_2) A^2 R_k, \quad A^2 R_k = R_{k-2} - 2R_{k-1} + R_k, \quad R_{-1} = R_{-2} = 0$$

where

$$\sum_0^{n-2} R_k x^k = \prod_{i=3}^n (q_i + p_i x).$$

It follows that

$$\sum_{j=0}^k \sum_{i=0}^j (b'_i - b_i) = (p'_1 p'_2 - p_1 p_2) R_k$$

which is nonnegative since R_k clearly satisfies $R_k \geq 0$ while

$$p'_1 p'_2 - p_1 p_2 = t(1 - t)(p_1^2 + p_2^2 - 2p_1 p_2) \geq 0.$$

(3) Consider an infinite sequence of independent binomial trials with probabilities of success

$$p_1, p_2, p_3, \dots, \quad 0 < p_i \leq 1.$$

Suppose $\sum_{i=1}^\infty p_i = \lambda < \infty$. The probabilities a_k of k successes can be calculated from the generating function

$$\sum_{k=0}^{\infty} a_k x^k = \prod_{i=1}^{\infty} (q_i + p_i x), \quad q_i + p_i = 1.$$

We assert that

$$(11) \quad \sum_{k=0}^{\infty} a_k \varphi(k) \leq \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \varphi(k)$$

for any convex sequence $\varphi(k)$ which grows as $k \rightarrow \infty$ no faster than an exponential. Combining the facts that $\log(k! a_k)$ is strictly concave (see [2], page 51) and that the 0th and 1st moment of

$$b_k = \frac{\lambda^k}{k!} e^{-\lambda} - a_k$$

vanish, we deduce, that $\{b_k\}_{k=0}^{\infty}$ changes sign exactly twice. Moreover,

$$a_0 = \prod_{i=1}^{\infty} (1 - p_i) < e^{-\sum p_i} = e^{-\lambda}$$

shows that $b_0 > 0$. Applying Lemma b gives (11). The same result can also be achieved by a suitable limiting process from Example 2.

This can also be written:

$$E_X(\varphi) \leq E_Y(\varphi)$$

where X is the number of successes in an infinite sequence of independent binomial trials and Y a suitable Poisson variable.

(4) Define a_k by means of the generating function

$$\sum_{k=0}^{\infty} a_k x^k = \prod_{i=1}^n \frac{q_i}{1 - p_i x}$$

where p_1, p_2, \dots, p_n are prescribed such that $0 < p_i < 1, p_i + q_i = 1$. We know

$$\sum_{k=0}^{\infty} \binom{n+k-1}{n-1} q^n p^k x^k = \left(\frac{q}{1 - px} \right)^n.$$

According to [2, p. 164],

$$\log \left[\frac{a_k}{\binom{n+k-1}{n-1} q^n p^k} \right]$$

is strictly convex as a sequence in k unless $p_1 = p_2 = \dots = p_n$.

In particular, if p is determined by the equation

$$\frac{n}{p} = \frac{1}{p_1} + \frac{1}{p_2} + \dots + \frac{1}{p_n},$$

and $p + q = 1$, then the hypotheses of Lemma b are satisfied, and we conclude that

$$(12) \quad \sum_{k=0}^{\infty} a_k \varphi(k) \geq \sum_{k=0}^{\infty} \binom{n+k-1}{n-1} p^k q^n \varphi(k)$$

for every convex sequence $\varphi(k)$. (Notice the reversal of the sense of inequality compared to (11).)

We may extend this as follows: If

$$\frac{p'_i}{q'_i} = \sum_{j=1}^n \pi_{ij} \frac{p_j}{q_j} \quad (i = 1, \dots, n) \quad 0 \leq p'_i \leq 1, \quad p'_i + q'_i = 1$$

and π_{ij} is again doubly stochastic then

$$(12') \quad \sum_{k=0}^{\infty} \varphi(k) a_k \geq \sum_{k=0}^{\infty} \varphi(k) a'_k$$

where φ is convex. (This includes the possibility that both sides are infinite.)

(5) The discrete version of Example 2 of § 1 asserts $\sum_0^{\infty} C_n \varphi(n) \leq (1 - \rho) \sum_{n=0}^{\infty} \rho^n \varphi(n)$ for all convex $\varphi(n)$ where $(\rho/1 - \rho) = \sum_0^{\infty} n C_n$ and $\{C_n\}$ satisfies $\sum_0^{\infty} C_n = 1$ and is a Pólya frequency sequence, i.e., $\det_{i,j=1,2} (C_{n_i - m_j}) \geq 0$ for $n_1 < n_2, m_1 < m_2$.

It is worthwhile to point out the connection of the conditions (i)–(iii) with the familiar condition of Karamata [2 Chap. 1] in order that

$$(13) \quad \sum_{i=1}^n \varphi(a_i) \leq \sum_{i=1}^n \varphi(a'_i)$$

whenever φ is convex. It is established in [2] that (13) holds when $\{a_i\}$ and $\{a'_i\}$ are arranged so that $0 \leq a_1 \leq a_2 \leq \dots \leq a_n, 0 \leq a'_1 \leq a'_2 \leq \dots \leq a'_n$ if and only if

$$(K_1) \quad \sum_{i=1}^r a_i \leq \sum_{i=1}^r a'_i \quad r = 1, 2, \dots, n - 1$$

$$(K_2) \quad \sum_{i=1}^n a_i = \sum_{i=1}^n a'_i .$$

Let $d\mu_1$ ($d\mu_2$) denote the measure with unit mass located at each of the points $\{a_i\}_{i=1}^n$ ($\{a'_i\}$). (If equalities are present among the a_i then we assign at the common point a mass equal to its multiplicity.)

We define

$$\nu_1(x) = \int_0^x d\mu_1 = \text{Number of } a_i < x$$

$$\nu_2(x) = \int_0^x d\mu_2 = \text{Number of } a'_i < x$$

for all $x > 0$. The relations (i)–(iii) become

$$(i)'' \quad \int d\mu_1(x) = \int d\mu_2(x) = n$$

$$(ii)'' \quad \int x d\mu_1(x) = \sum_{i=1}^n a_i = \sum_{i=1}^n a'_i = \int x d\mu_2(x)$$

$$(iii)'' \quad \int_0^x \nu_1(y) dy \geq \int_0^x \nu_2(y) dy .$$

It is not difficult to verify directly the equivalence of (iii)'' and (K₁) which we do below. (In view of this the characterization by means of (i)''–(iii)'' can be regarded as a generalization of the Karamata theorem pertaining to the inequalities (K₁.) Note that by its definition, $\nu_1(y) = k$ in $(a_k, a_{k+1}]$ and $\nu_2(y) = k$ in $(a'_k, a'_{k+1}]$.

We begin by observing that

$$\begin{aligned} \int_0^{a_k} \nu_1(y) dy &= (a_2 - a_1) + 2(a_3 - a_2) + \dots + (k - 1)(a_k - a_{k-1}) \\ &= -a_1 - a_2 - \dots - a_k + ka_k . \end{aligned}$$

On the other hand if $l = l(k)$ is chosen so that $a'_i \leq a_k < a'_{l+1}$, then

$$\int_0^{a_k} \nu_2(y) dy = -a'_1 - a'_2 - \dots - a'_l + la_k .$$

If we now assume (iii)'' for all x , then it holds in particular for $x = a_k$, and we have

$$(14) \quad -\sum_1^k a_i + ka_k \geq -\sum_1^l a'_i + la_k \quad \text{or} \quad \sum_1^k a_i + la_k \leq \sum_1^l a'_i + ka_k .$$

From (14) we now deduce (K₁):

If $l \leq k$, then $\sum_1^l a'_i + la_k \leq \sum_1^k a'_i + la_k$ since $a'_{l+1} \geq a_k$ while if $l > k$, $\sum_1^l a'_i \leq \sum_1^k a'_i + (l - k)a_k$. In either case (14) implies (K₁).

On the other hand if (14) holds for $k = 1, \dots, n$, then (iii)'' will hold for $x = a_1, a_2, \dots, a_n$. Since $\int_0^x \nu_1(y) dy$ is linear in the intervals between these points, while $\int_0^x \nu_2(y) dy$ is convex, the inequality (iii)'' is still true between these points, i.e., for all x . Thus it suffices to show (K₁) implies (14) to show that (K₁) implies (iii)'' and thus complete the proof of equivalence. Assume therefore $\sum_1^k a_i \leq \sum_1^k a'_i$ for $k = 1, \dots, n$; if $l \leq k$, $\sum_1^k a_i \leq \sum_1^l a'_i + (k - l)a_k$ while if $l > k$ we have $\sum_1^k a'_i + (l - k)a_k \leq \sum_1^l a'_i$. Thus $\sum_1^k a_i \leq \sum_1^k a'_i \leq \sum_1^l a'_i + (k - l)a_k$, and so in every case we have the inequality between the extremes, which is precisely (14).

3. Generalized convexity inequalities. Our objective in this section is to characterize the duals to cones which generalize the cone defined by (2) of § 1. We formulate and analyze a continuous version in detail; corresponding results in the discrete case are actually simpler to discuss.

To consider the simplest continuous analogy, let $u_1(x), \dots, u_k(x)$ be positive C^k functions on the finite interval $[a, b]$, and let

$$\begin{aligned}
 \psi_1(x) &= u_1(x) , \\
 \psi_2(x) &= u_1(x) \int_a^x u_2(\xi_1) d\xi_1 , \\
 &\dots\dots\dots \\
 \psi_k(x) &= u_1(x) \int_a^x u_2(\xi_1) \dots \int_a^{\xi_{k-2}} u_k(\xi_{k-1}) d\xi_{k-1} \dots d\xi_1 .
 \end{aligned}
 \tag{15}$$

If we let D_j be the differential operator of first order defined by

$$D_j f = \frac{d}{dx} \frac{1}{u_j} f$$

then clearly

$$D_j \dots D_1 \psi_{j+1} = u_{j+1} > 0 , \qquad j = 1, \dots, k - 1$$

and

$$D_j \dots D_1 \psi_j = 0 , \qquad j = 1, \dots, k$$

and so

$$D_k \dots D_1 \psi_j = 0 , \qquad j = 1, \dots, k .$$

The Wronskian $W(\psi_1, \dots, \psi_j) = u_1^j u_2^{j-1} \dots u_j$ and so is positive in $[a, b]$ for $j = 1, \dots, k$, so that the differential operator $D_k \dots D_1$ enjoys "property W " of [10]. In particular, no nontrivial linear combination of ψ_1, \dots, ψ_j can have more than $j - 1$ zeros in $[a, b]$ (counting multiplicities) and ψ_1, \dots, ψ_k form a basis for the solutions of $D_k \dots D_1 \psi = 0$. The cone we shall be concerned with is closely related to the totality of C^k functions φ satisfying

$$D_k \dots D_1 \varphi \geq 0$$

namely the cone of all φ satisfying

$$\begin{aligned}
 (18) \quad & \left| \begin{array}{ccc} \psi_1(x_1) & \dots & \psi_k(x_1) & \varphi(x_1) \\ \vdots & & & \\ \psi_1(x_{k+1}) & \dots & \psi_k(x_{k+1}) & \varphi(x_{k+1}) \end{array} \right| \geq 0 \\
 & \text{for } a \leq x_1 < \dots < x_{k+1} \leq b .
 \end{aligned}$$

This cone we denote $C(\psi_1, \dots, \psi_k; [a, b])$.

Between (17) and (18) the following relations are known to hold [5]: (17) implies (18), and strictness in (17) implies strictness in (18). If (18) holds and φ is assumed C^k , then (17) is true; strictness of (18) need not imply strictness in (17). In addition (18) itself imposes some smoothness restrictions on φ (see [5]).

Note that $D_{j-1} \dots D_1 \psi_j = u_j > 0$ which is a case of strictness in (17), necessarily implies

$$(17') \quad \bar{\psi} \begin{pmatrix} x_1, x_2, \dots, x_j \\ 1, 2, \dots, j \end{pmatrix} = \begin{vmatrix} \psi_1(x_1) \dots \psi_j(x_1) \\ \vdots \\ \psi_1(x_j) \dots \psi_j(x_j) \end{vmatrix} > 0$$

for $j = 1, \dots, k$, i.e., $\psi_j \in C(\psi_1, \dots, \psi_{j-1}; [a, b])$. In the case $u_1(x) = u_2(x) \equiv 1$, or $\psi_1(x) \equiv 1, \psi_2(x) \equiv x$ which was discussed in § 1, we see that (17) reduces to $d^2\varphi/dx^2 \geq 0$, while (18) is the definition of convexity.

It is possible to remove some of the restrictions we have imposed on ψ_1, \dots, ψ_k and thus to generalize even the class of cones considered here. For example in § 5 we weaken the assumption (17). The restriction to finite $[a, b]$ is entirely inessential, as in § 1, and the modifications and reinterpretations of the treatment below we leave to the reader. We shall feel free to give examples from the infinite as well as the finite case at the end of this section. It is also possible to remove the assumption that ψ_1, \dots, ψ_k be C^k , merely requiring (17'). We will not needlessly burden the exposition by going into the corresponding analogues to (15) and (16) and the corresponding reformulation of Theorem 1 here.

The functions of $C(\psi_1, \dots, \psi_k; [a, b])$ have been called "generalized convex functions in $[a, b]$ "; the domain $[a, b]$ is omitted without ambiguity whenever convenient. This terminology was employed in [7].

We are interested in characterizing the dual cone to $C(\psi_1, \dots, \psi_k; [a, b])$, i.e. the set of signed measures $d\mu$ on $[a, b]$ such that

$$(19) \quad \int_a^b \varphi(x) d\mu(x) \geq 0$$

for all $\varphi \in C(\psi_1, \dots, \psi_k)$. For simplicity assume first that $d\mu(x) = f(x)dx$. The extension to the general case will be clear. We remark that every φ in $C(\psi_1, \dots, \psi_k)$ can be approximated weakly by solutions of (17), so that it suffices to assume in (19) that $\varphi \in C^k$.

LEMMA. *The "moment conditions"*

$$(a) \quad \int_a^b f(x) \psi_j(x) dx = 0 \quad j = 1, \dots, k$$

are equivalent to the conditions

$$(b) \quad I_1(f)(b) = 0, \quad I_2 I_1(f)(b) = 0, \quad \dots, \quad (I_k \dots I_1)(f)(b) = 0$$

where the integral operators $I_j, j = 1, \dots, k$ are defined by

$$I_j(d\mu)(x) = -\int_a^x u_j(t) d\mu(t)$$

and

$$I_j(f)(x) = -\int_a^x u_j(t) f(t) dt$$

when applied to measures and functions respectively.

Proof. We proceed by induction on k . It is only necessary to observe that the functions and operators

$$\frac{d}{dx} \left(\frac{\psi_2}{\psi_1} \right), \dots, \frac{d}{dx} \left(\frac{\psi_k}{\psi_1} \right), I_1 f(x)$$

and I_2, \dots, I_k satisfy the induction hypotheses and play the roles of $\psi_1, \dots, \psi_k, f, I_1, \dots, I_k$, respectively. We omit the formal details.

We may now state the main theorem of this section.

THEOREM 1. *A necessary and sufficient condition for $\int_a^b f \varphi dx \geq 0$ whenever φ satisfies (17) is that f satisfy the "moment conditions" (a) and also*

$$(20) \quad (I_k \dots I_1) f(x) \geq 0 \quad a \leq x \leq b.$$

Proof. Clearly conditions (b) are necessary, since $\pm \psi_j$ satisfy (17) for $j = 1, \dots, k$, and by the lemma, this establishes the necessity of (a) as well. However, when (a) holds integration by parts shows that

$$\int_a^b f \varphi dx = \int_a^b (D_k \dots D_1 \varphi)(I_k \dots I_1 f) dx.$$

Thus (20) is sufficient in the presence of (a). Since we may take $D_k \dots D_1 \varphi$ to be an arbitrary positive continuous function in $[a, b]$, we see that (20) is also necessary.

The same proof shows that a necessary and sufficient condition on a measure $d\mu$ in order that $\int_a^b \varphi d\mu \geq 0$ whenever φ is in $C(\psi_1, \dots, \psi_k; [a, b])$ is that

$$\int_a^b \psi_j d\mu = 0 \quad j = 1, \dots, k$$

and

$$(I_k \dots I_1) d\mu(x) \geq 0 \quad a \leq x \leq b.$$

We make the standard abuse of terminology and say that $f(x)$ is in the dual cone if the measure $f(x)dx$ is in the dual cone.

We now establish a lemma (Lemma B below) which is useful in providing a sufficient condition that $\int_a^b f\varphi dx \geq 0$ for all φ satisfying $D_k \cdots D_1\varphi \geq 0$. We begin with a preliminary lemma.

LEMMA A. *If $\Psi \begin{pmatrix} x_1 \cdots x_j & x \\ 1 \cdots j & j+1 \end{pmatrix} > 0$ for $a \leq x_1 < \cdots < x_j \leq b$ (see (17')), $j = 1, \dots, k$ and $\int_a^b f(x)\psi_j(x)dx = 0$, $j = 1, \dots, k$ and f is continuous and nonnull, then f must have at least k sign-changes.*

This is well-known, but we give the proof for the sake of completeness. If there were j points $a < x_1 < \cdots < x_j < b$ such that f is of constant sign and nonnull in each of $[a, x_1], [x_1, x_2], \dots, [x_j, b]$, the signs alternating between successive intervals, where $j \leq k - 1$, then the function

$$\begin{aligned} \Psi(x) &= \Psi \begin{pmatrix} x_1 \cdots x_j & x \\ 1 \cdots j & j+1 \end{pmatrix} \\ &= \begin{vmatrix} \psi_1(x_1) \cdots \psi_{j+1}(x_1) \\ \vdots \\ \psi_1(x_j) \cdots \psi_{j+1}(x_j) \\ \psi_1(x) \cdots \psi_{j+1}(x) \end{vmatrix} = C_1\psi_1(x) + \cdots + C_{j+1}\psi_{j+1}(x) \quad (C_{j+1} > 0), \end{aligned}$$

would enjoy the same alternation property, vanish only at $x = x_1, \dots, x_j$, so that $f(x) \cdot \Psi(x)$ would be nonvanishing and of constant sign, contradicting $\int_a^b f(x)\psi_i(x)dx = 0$ ($i = 1, 2, \dots, k$).

LEMMA B. *If, under the same assumptions as Lemma A, f changes sign exactly k times in $[a, b]$, and is nonnegative and nonnull in some one-sided neighborhood of b then $\int_a^b f\varphi dx \geq 0$ for all $\varphi \in C(\psi_1, \dots, \psi_k)$.*

Proof. Let $[a, x_1], \dots, [x_k, b]$ be intervals of alternating sign for f , and let

$$\begin{aligned} \Psi(x) &= \begin{vmatrix} \varphi(x) & \psi_1(x) & \cdots & \psi_k(x) \\ \varphi(x_1) & \psi_1(x_1) & \cdots & \psi_k(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(x_k) & \psi_1(x_k) & \cdots & \psi_k(x_k) \end{vmatrix} \\ &= \Psi \begin{pmatrix} x_1 \cdots x_k \\ 1 \cdots k \end{pmatrix} \varphi(x) + \text{linear combinations of } \psi_1(x), \dots, \psi_k(x). \end{aligned}$$

Then

$$\int_a^b f \Phi dx = \Psi \begin{pmatrix} x_1 \cdots x_k \\ 1 \cdots k \end{pmatrix} \int_a^b f \varphi dx$$

where $f\Phi$ is nonnegative throughout the interval and nonnull in some neighborhood of b .

We may also give a proof of Lemma B by repeated application of the lemma that if $g(x) = \int_a^x u(t)f(t)dt$, $u(t) > 0$ in $[a, b]$ and $g(b) = 0$, then the number of zeros of g is less than the number of zeros of f by at least 1. Applying this k times for u_1, \dots, u_k , we establish that the minimal sign change condition implies $I_k \cdots I_1 f$ is of constant sign. Any further information concerning f to assure that the sign of $I_k \cdots I_1 f$ is indeed positive then constitutes a sufficient condition to assure that f is in the dual cone to $C(\psi_1, \dots, \psi_k; [a, b])$.

We remark that Lemmas A and B remain valid for measures $d\mu(x)$ in place of $f(x)dx$, and in particular for discrete versions. The appropriate notion of sign change is described by subdivision of $[a, b]$ into consecutive intervals in each of which $d\mu$ is of alternating sign and not null. We leave the details to the reader.

EXAMPLES.

(1) The first example we give concerns the cone $C(1)$ (i.e., $k = 1$, $\psi_1(x) \equiv 1$). These are the functions φ satisfying

$$\begin{vmatrix} 1 & \varphi(x_1) \\ 1 & \varphi(x_2) \end{vmatrix} \geq 0$$

for $x_1 \leq x_2$, i.e., the increasing functions. We cite the inequality of Steffanson [1]: if $0 \leq g(t) \leq 1$ in $[a, b]$, and $\varphi \in C(1; [a, b])$ then

$$(S) \quad \int_a^{a+c} \varphi(t) dt \leq \int_a^b \varphi(t)g(t) dt \leq \int_{b-c}^b \varphi(t) dt$$

where

$$c = \int_a^b g(t) dt .$$

We prove the first inequality. Let $g_c(t)$ be the characteristic function of the interval $[a, a+c]$. Then $g(t) - g_c(t)$ satisfies the moment condition with respect to $\psi_1(x) \equiv 1$ by the choice of c , is negative near a , and has exactly one change of sign (namely at $a+c$). Thus Lemma B applies. The second inequality is proved similarly.

(2) The discrete version of the cone $C(1)$ consists of the cone of increasing sequences. The corresponding necessary and sufficient condition that $\sum_{i=1}^n a_k d_k \geq 0$ for all increasing $\{a_k\}$ is that $\sum_{i=1}^n d_k = 0$ and

$\sum_1^k d_j \geq 0, k = 1, \dots, n$. An application of this yields the well-known theorem of Tchbeycheff on rearrangements. Let $a_1 \leq \dots \leq a_n$, and let σ denote a permutation of $\{1, \dots, n\}$. Then

$$\sum_1^n a_k d_{\sigma(k)} \leq \sum_1^n a_k d_{\sigma_0(k)}$$

where σ_0 is that permutation such that

$$d_{\sigma_0(1)} \leq d_{\sigma_0(2)} \leq \dots \leq d_{\sigma_0(n)} .$$

(3) Another application concerning the same cone as the previous examples is the following theorem of Mirsky [9]. Let A and B be real symmetric matrices with eigenvalues $\alpha_1 \leq \dots \leq \alpha_n, \beta_1 \leq \dots \leq \beta_n$, respectively. Then $\text{trace } AB \leq \sum \alpha_i \beta_i$. Clearly, there is no loss of generality in choosing a basis so that B is in diagonal form

$$\begin{pmatrix} \beta_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \beta_n \end{pmatrix}$$

in which case $\text{trace } AB = \sum_1^n a_{ii} \beta_i$. Since $\sum_1^n a_{ii} = \sum_1^n \alpha_i = \text{trace } A$ the necessary moment condition is satisfied by the sequence $\{(\alpha_i - a_{ii})\}$. It remains to verify

$$\sum_{i=1}^k \alpha_i \leq \sum_{i=1}^k a_{ii} \qquad k = 1, \dots, n .$$

We will employ our earlier examples to further illustrate the utility of our methods. In virtue of the Karamata result of § 2 it suffices to establish

$$\sum_1^n \varphi(\alpha_i) \geq \sum_{i=1}^n \varphi(a_{ii})$$

for arbitrary convex φ . We remark that the left-hand side is $\text{trace } \varphi(A)$. Let $A = \int \lambda dE_\lambda$ be the spectral decomposition of A . Then if x is a vector such that $(x, x) = 1$ then $(dE_\lambda x, x)$ is nonnegative measure of total mass 1 on the real axis. By Jensen's inequality, we have

$$\varphi((Ax, x)) = \varphi\left(\int \lambda (dE_\lambda x, x)\right) \leq \int \varphi(\lambda) (dE_\lambda x, x) = (\varphi(A)x, x) .$$

In particular taking for x the basis vectors with 1 in the i th component and other components zero and adding the resulting inequalities gives the desired result.

(4) Berwald's inequality [1]. This is a generalization of Favard's inequality (Example 1 of § 1) using the cone $C(1, \psi(x))$ (where $\psi(x)$

is necessarily increasing (cf. (17')) in place of $C(1, x)$. Let $f(x)$ be concave, or more generally such that $d\mu_f$ is increasing in the notation of § 1, Example 1. Let \bar{z} be the sole positive root of

$$\frac{1}{\bar{z}} \int_0^{\bar{z}} \psi(y) dy = \frac{1}{b-a} \int_a^b \psi(f(x)) dx = \int_0^\infty \psi(y) d\mu_f(y).$$

The result of Berwald is that

$$(B) \quad \frac{1}{\bar{z}} \int_0^{\bar{z}} \varphi(y) dy \geq \frac{1}{b-a} \int_a^b \varphi(f(x)) dx$$

for $\varphi \in C(1, \psi)$. The proof is identical with that of (F), namely an application of Lemma B. The definition of \bar{z} is precisely the moment condition for the corresponding measure $d\nu(y) - d\mu_f(y)$ ($d\nu(y) = 1/\bar{z}$ for $0 \leq y \leq \bar{z}$ and equals zero elsewhere) with respect to the function $\psi(y)$.

(5) Let F and G be probability distribution functions of positive random variables with the same mean, so that we have

$$(a) \quad \int_0^\infty dF = \int_0^\infty dG = 1$$

$$(b) \quad \int_0^\infty (1 - F) dx = \int_0^\infty (1 - G) dx = \mu.$$

Assume further that

$$(c) \quad 1 - F = e^{-\psi_1(x)}, \quad 1 - G = e^{-\psi_2(x)}$$

where ψ_2 is in $C(\psi_1)$ and ψ_1 is strictly increasing. (This is equivalent to $(1/y)\psi_2(\psi_1^{-1}(y))$ being in $C(1)$, i.e., increasing.)

Such distribution functions have been studied in [6], especially under the additional assumption of convexity for ψ_1 and ψ_2 .

We assert that (a), (b), and (c) together imply

$$\int_0^\infty \varphi dF \leq \int_0^\infty \varphi dG$$

for convex φ . For simplicity we will assume φ is in C^1 , since the general case follows by the approximation procedure mentioned in § 1. Moreover, by (a) we may without loss of generality take $\varphi(0) = 0$. Note that

$$\int_0^\infty \varphi dF = \int_0^\infty \varphi'(1 - F) dx = \int_0^\infty \varphi' e^{-\psi_1} dx$$

and so we need to establish

$$\int_0^\infty \varphi'(e^{-\psi_1} - e^{-\psi_2}) dx \leq 0$$

for increasing φ' , i.e., that $e^{-\psi_2} - e^{-\psi_1}$ is in the dual cone to $C(1)$. We note that

$$\text{sign}(e^{-\psi_2} - e^{-\psi_1}) = \text{sign}(\psi_1(x) - \psi_2(x)) = \text{sign}\left(1 - \frac{\psi_2(\psi_1^{-1}(y))}{y}\right)$$

where $y = \psi_1(x) > 0$ for $x > 0$. By hypothesis (b) we know that the difference must have at least one strict sign change (unless identically 0), and Lemma B applies.

We note that an arbitrary strictly decreasing function $g(y)$ could have been used in the hypothesis (c)

$$1 - F = g(\psi_1(x)), \quad 1 - G = g(\psi_2(y))$$

providing $0 < g(y) < 1$ implies $y > 0$.

(6) Suppose now that the assumptions on ψ_1 and ψ_2 are that ψ_1 is strictly increasing and ψ_2 convex with respect to ψ_1 , that is, ψ_2 is in $C(1, \psi_1)$ (or equivalently, $\psi_2(\psi_1^{-1}(u))$ is convex in u). Thus

$$\begin{vmatrix} 1 & \psi_1(x_1) & \psi_2(x_1) \\ 1 & \psi_1(x_2) & \psi_2(x_2) \\ 1 & \psi_1(x_3) & \psi_2(x_3) \end{vmatrix} \geq 0 \quad \text{for } x_1 < x_2 < x_3.$$

We assume $F(0) = G(0) = 0$ so that $\psi_1(0) = \psi_2(0) = 0$, as well as (a) and (b) of Example 5. We have

$$e^{-\psi_2(x)} - e^{-\psi_1(x)} = e^{-\psi(u)} - e^{-u}$$

where $\psi(u) = \psi_2(\psi_1^{-1}(u))$ is convex by assumption, and $\psi(0) = 0$. Now we see that $\psi(u) - u$ can have at most two intervals of positivity, but because the difference vanishes at $u = 0$, it must in fact be negative near 0, and ultimately [positive (the possibility of $\psi(u) \geq u$ throughout $(0, \infty)$ is excluded by Lemma A). Thus we can say in this case also for φ convex, that

$$\int_0^\infty \varphi dF \leq \int_0^\infty \varphi dG.$$

We remark that the hypotheses $\psi_2 \in C(\psi_1)$ and $\psi_2 \in C(1, \psi_1)$ are not unrelated (cf. [2], Th. 127). We defer a complete discussion of such relations to a later paper.

4. Convolution of inequalities. In this section we concentrate

on the fact that the cone $C = C(1, x, \dots, x^{k-1}; (-\infty, +\infty))$ is invariant under translation. That is, if we introduce the translation operator $T_y\varphi(x) = \varphi(x + y)$ then the cone is mapped into itself by T_y . (We note that one can introduce more generally a translation operator T_y associated with the cone $C(\psi_1, \dots, \psi_k)$ when the cone is defined as in § 3 in terms of the functions $u_1(x), \dots, u_k(x)$. This T_y will, in particular, reduce to ordinary translation when each $u_i(x)$ is constant. The precise definition of T_y and the corresponding modifications of the remainder of this section will not be given here, but will be discussed elsewhere.)

Turning our attention now to the dual cone C^* , we observe that

$$\int_{-\infty}^{+\infty} f(x)(T_y\varphi)(x)dx = \int_{-\infty}^{+\infty} (T_{-y}f)(x)\varphi(x)dx$$

i.e., T_y induces the adjoint operator T_y^* which in this case is merely T_{-y} . It follows that the dual cone is itself invariant with respect to translation. An immediate consequence is that for all φ in $C(1, \dots, x^{k-1}; (-\infty, +\infty))$ and f in the dual cone, we have

$$(21) \quad \int_{-\infty}^{+\infty} \hat{f}(x)\varphi(x)dx \geq 0$$

whenever

$$(22) \quad \hat{f}(x) = \int_{-\infty}^{+\infty} (T_{-y}f)(x)d\sigma(y)$$

for any nonnegative measure $d\sigma$. This is to be interpreted as allowing the value $+\infty$ for $\int_{-\infty}^{+\infty} \hat{f}(x)\varphi(x)dx$ in the absence of further restrictions on the measure $d\sigma$, but we shall give some examples, in the case $k = 2$, where convergence of the integral is apparent. It is also clear that the above considerations extend to

$$(23) \quad d\hat{\mu}(x) = d\int_{-\infty}^{+\infty} (T_{-y}\mu)(x)d\sigma(y)$$

where the measure $d\mu$ is in the dual of $C(1, x, \dots, x^{k-1}; (-\infty, +\infty))$. In particular, if $f_1(x) - f_2(x) = h_1(x) \in C^*$, $g_1(x) - g_2(x) = h_2(x) \in C^*$, where C and hence C^* is a cone invariant under translation, then

$$f_1 * g_1 - f_2 * g_2 = h_1 * g_2 + f_2 * h_2 + h_1 * h_2$$

is again in C^* . Let us introduce the notation $d\mu_1 \ll d\mu_2$ to denote $d\mu_2 - d\mu_1 \in C^*$. Then we have in general

$$(24) \quad d\mu_1^{(1)} * \dots * d\mu_1^{(n)} \ll d\mu_2^{(1)} * \dots * d\mu_2^{(n)}$$

where $d\mu_1^{(i)} \ll d\mu_2^{(i)}$, $i = 1, \dots, n$. If $d\mu_1$ and $d\mu_2$ are the probability distributions of random variable X_1 and X_2 , respectively, we write $X_1 \ll X_2$ to indicate $E_{X_1}(\varphi) \leq E_{X_2}(\varphi)$ for all φ in C or, equivalently $d\mu_2 - d\mu_1 \in C^*$. In this notation we have $X_i \ll Y_i$, $i = 1, \dots, n$, implies

$$(25) \quad X_1 + \dots + X_n \ll Y_1 + \dots + Y_n$$

EXAMPLES.

(1) If G is the distribution function of a random variable Y of mean 0, and F the distribution function of a random variable X of finite mean, then

$$dF \ll (dF) * (dG) \quad \text{or} \quad X \ll X + Y.$$

This results from convoluting both sides of

$$dG_0 \ll dG$$

with dF , where

$$G_0(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

(cf. Jensen's inequality, (5)). Generally, $X + (\text{mean } Y) \ll X + Y$.

(2) Recall that $f_i(x) \ll p(x; \lambda_i)$, $i = 1, \dots, n$ where $f_i(x)$ is a Pólya frequency density of order two on $[0, \infty]$ with mean λ_i and

$$(26) \quad p(x, \lambda) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(cf. § 1, Example 2). It follows that

$$(27) \quad f_1 * \dots * f_n \ll p(x; \lambda_1) * \dots * p(x, \lambda_n).$$

If we denote the n -fold convolution of the Pólya frequency density $f(x)$ with itself by $f^{(n)}$ then from $f(x) \ll p(x, \lambda)$ we deduce

$$(28) \quad f^{(n)}(x) \ll p^{(n)}(x, \lambda)$$

and, of course,

$$(29) \quad p^{(n)}(x, \lambda) = \begin{cases} \frac{x^{n-1}}{\lambda^n \Gamma(n)} e^{-x/\lambda} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

It is observed that $f^{(n)}(x)$ is again in Pólya frequency density of

order 2 with mean $n\lambda$ and null on $[-\infty, 0]$ from which we obtain

$$(30) \quad f^{(n)}(x) \ll p(x, n\lambda).$$

If we take for f in (30) the choice of $f(x) = p(x, \lambda)$, we see that

$$(31) \quad p^{(n)}(x, \lambda) \ll p(x, n\lambda).$$

Comparing (28), (30) and (31), shows that (28) is sharper than (30).

In the previous example we have restricted ourselves to the consideration of densities which are PF_2 , and have used the fact that the class of such densities is closed under convolution. We now exploit the slightly deeper fact due to Barlow and Marshall that if F_1 and F_2 are distribution functions such that $1 - F_1$ and $1 - F_2$ are PF_2 , then $1 - F_3$ is again PF_2 where

$$F_3(x) = \int F_1(x - y)dF_2(y) \quad (\text{see ref. [5]}).$$

Assume therefore that

$$1 - F_1 = e^{-\psi_1} \quad \text{and} \quad 1 - F_2 = e^{-\psi_2}$$

where ψ_1 and ψ_2 are convex increasing. We know by Example 5 of § 3 that

$$1 - F_i(x) \ll e^{-x/\lambda_i} \quad i = 1, 2$$

where λ_i is the mean of the distribution $F_i(x)$. Following the method of Example 2, we obtain

$$1 - F_3 \ll e^{-x/\lambda_1} * e^{-x/\lambda_2}$$

and

$$1 - F_1 * \dots * F_n \ll e^{-x/\lambda_1} * \dots * e^{-x/\lambda_n}.$$

As an example of this, we obtain

$$(32) \quad 1 - F^{(n)} \ll \frac{x^{n-1}}{\Gamma(n)\lambda^n} e^{-x/\lambda}$$

where $F(x)$ is a PF_2 cumulant of mean λ .

This is to be compared with the simpler relation

$$(33) \quad 1 - F^{(n)} \ll \frac{1}{n\lambda} e^{-x/n\lambda}.$$

which is weaker than (32).

(3) Consider densities f on $(0, \infty)$ possessing all moments, with mean μ , and let $f^{(n)}$ denote the n -fold convolution of f with itself.

Let g be continuous and bounded (or of growth no greater than exponential).

Let

$$B_n(\lambda, g) = \int_0^\infty g\left(\frac{x\lambda}{n\mu}\right) f^{(n)}(x) dx .$$

Then the following facts about $B_n(\lambda, g)$ are established in [7]

(a) $B_n(\lambda, 1) = 1$

(b) $B_n(\lambda, x) = \lambda$

(c) If g is a polynomial in x , $B_n(\lambda, g)$ is a polynomial of the same degree in λ . ((a) and (b) are special instances of this.)

(d) If g is bounded and continuous then $\lim_{n \rightarrow \infty} B_n(\lambda, g) = g(\lambda)$ uniformly on bounded intervals. This is actually a simple application of the law of large numbers, and in fact one can assert more generally, the same conclusion to hold if $f_1 * \dots * f_n$ replaces $f^{(n)}$, providing the densities f_1, \dots, f_n have variances $\sigma_1^2, \dots, \sigma_n^2$ permitting the application of the law of large numbers.

We remark also that the class of summability or approximating kernels obtained in this way contains, as a limiting case of a discrete analogue, the celebrated Bernstein polynormal kernels.

(e) If g is convex, then $B_n(\lambda, g)$ is convex in λ and

(f) $B_n(\lambda, g) \geq g(\lambda)$.

Property (e) requires only the positivity of f and property (f) uses the fact that the mapping $g \rightarrow B_n(\lambda, g)$ preserves convexity, positivity, and linear functions.

To these results, quoted from [7], we now add an application of our main theorem.

THEOREM. *Whenever f is PF_∞ , vanishes for $x < 0$, and g is convex, then $B_n(g, \lambda)$ is monotone in n .*

Proof. We first consider the case $f(x) = (1/\mu) e^{-x/\mu}$ in $[0, \infty]$, null elsewhere, which clearly satisfies

$$\int_0^\infty f(x) dx = 1, \quad \int_0^\infty x f(x) dx = \mu .$$

We desire to show that

$$\int_0^\infty g\left(\frac{x\lambda}{n\mu}\right) f^{(n)}(x) dx \geq \int_0^\infty g\left(\frac{x\lambda}{(n+1)\mu}\right) f^{(n+1)}(x) dx .$$

Let $x = n\xi$ in the left hand side and $x = (n+1)\xi$ on the right, so that the desired inequality is

$$(34) \quad \int_0^\infty g\left(\frac{\xi\lambda}{\mu}\right) \{n f^{(n)}(n\xi) - (n+1) f^{(n+1)}((n+1)\xi)\} d\xi \geq 0$$

for convex g . The presence of the scale factor λ being of no importance, it suffices to show that

$$nf^{(n)}(nx) - (n + 1)f^{(n+1)}((n + 1)x)$$

has exactly two changes of sign and is positive near 0. Since

$$\int_0^\infty nf^{(n)}(nx)dx = \int_0^\infty f^{(n)}(x)dx = 1$$

and

$$\int_0^\infty nxf^{(n)}(nx)dx = \frac{1}{n} \int_0^\infty xf^{(n)}(x)dx = \mu$$

we see that the difference has zero moments with respect to 1 and x and so has at least two sign changes. On the other hand, it is immediate from the expression

$$f^{(n)}(x) = \frac{1}{\Gamma(n)} \frac{x^{n-1}}{\mu^n} e^{-x/\mu} \quad (x > 0)$$

that the difference has at most two sign changes in $(0, \infty)$. The same result is true for

$$f_a(x) = \begin{cases} \frac{1}{\mu} e^{-(x-a)/\mu} & x > a \\ \mu & x = a \\ 0 & x < a \end{cases} \quad a > 0.$$

We now remark that if f and g are such that $nf^{(n)}(nx) \gg (n + 1)f^{(n+1)}((n + 1)x)$ and $ng^{(n)}(nx) \gg (n + 1)g^{(n+1)}((n + 1)x)$ then $nh^{(n)}(nx) \gg (n + 1)h^{(n+1)}(n + 1)x$ where $h = f * g$, by § 4. Indeed

$$nh^{(n)}(nx) = nf^{(n)}(nx) * ng^{(n)}(nx)$$

the left-hand side representing the density of $(1/n)\{(X_1 + Y_1) + \dots + (X_n + Y_n)\}$ where each X_i has density f and each Y_i has density g .

It follows from the general structure of PF_∞ densities on $(-\infty, +\infty)$ which vanish in $(-\infty, 0)$ as given in [11], that the inequality (34) is true for the entire class of PF_∞ distributions on the positive axis, as asserted.

5. An approximation procedure. The assumption that strict inequality in (17') prevails can be considerably weakened and a result of the type of Lemma B will hold. Specifically, assume ψ_j are continuous and that

$$(17'') \quad \psi \begin{pmatrix} x_1, x_2, \dots, x_j \\ 1, \dots, j \end{pmatrix} = \begin{vmatrix} \psi_1(x_1), & \psi_1(x_2), & \dots, & \psi_1(x_j) \\ \psi_2(x_1), & \psi_2(x_2), & \dots, & \psi_2(x_j) \\ \vdots & & & \\ \psi_j(x_1), & \psi_j(x_2), & \dots, & \psi_j(x_j) \end{vmatrix} \geq 0$$

$$(a \leq x_1 < x_2 < \dots < x_j \leq b)$$

and does not vanish identically ($j = 1, 2, \dots, k$). We associate with $\psi_j(x)$ the function

$$(35) \quad \tilde{\psi}_j(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left[-\frac{1}{2\sigma^2}(x-y)^2\right] \psi_j(y) dy$$

$$j = 1, 2, \dots, k$$

($\sigma > 0$) and with φ the function

$$\tilde{\varphi}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left[-\frac{1}{2\sigma^2}(x-y)^2\right] \varphi(y) dy .$$

If ψ_j satisfy (17'') and φ fulfills (18) then the functions $\tilde{\psi}_1, \tilde{\psi}_2, \dots, \tilde{\psi}_k, \tilde{\varphi}$ enjoy the inequalities (17') and (18). Now however, $\tilde{\psi}_1, \dots, \tilde{\psi}_k$ satisfy the requirements of Theorem 1. (Frequently even the assumption that ψ_1, \dots, ψ_k are continuous can be relaxed.) These facts are an easy consequence of the assumptions and the identity

$$\tilde{\Psi} \begin{pmatrix} x_1, x_2, \dots, x_j \\ 1, \dots, j \end{pmatrix} = \int_{a \leq y_1 < y_2 < \dots < y_j \leq b} \dots \int N_\sigma \begin{pmatrix} x_1, x_2, \dots, x_j \\ y_1, y_2, \dots, y_j \end{pmatrix} \times \tilde{\Psi} \begin{pmatrix} y_1, y_2, \dots, y_j \\ 1, 2, \dots, j \end{pmatrix} dy_1 dy_2 \dots dy_j$$

where

$$N_\sigma \begin{pmatrix} x_1, x_2, \dots, x_j \\ y_1, y_2, \dots, y_j \end{pmatrix} = \frac{1}{(\sqrt{2\pi}\sigma)^j} \begin{vmatrix} \exp\left(-\frac{1}{2\sigma^2}(x_1-y_1)^2\right) \dots \exp\left(-\frac{1}{2\sigma^2}(x_1-y_j)^2\right) \\ \vdots \\ \exp\left(-\frac{1}{2\sigma^2}(x_j-y_1)^2\right) \dots \exp\left(-\frac{1}{2\sigma^2}(x_j-y_j)^2\right) \end{vmatrix} > 0$$

for all

$$a \leq \left\{ \begin{matrix} x_1 < x_2 < \dots < x_j \\ y_1 < y_2 < \dots < y_j \end{matrix} \right\} \leq b$$

(see [5]). Clearly as $\sigma \rightarrow 0+$, $\tilde{\psi}_j(x) \rightarrow \psi_j(x)$ boundedly and $\tilde{\varphi}(x) \rightarrow \varphi(x)$ for $a < x < b$.

Assume now that $f(x) \in C^1$ in $[a, b]$, exhibits exactly k sign changes,

is orthogonal to each $\psi_i(x)$ and has only simple zeros. We define

$$\tilde{f}(x) = f(x) - \sum_{j=1}^k a_j(\sigma) \tilde{\psi}_j(x)$$

where $a_i(\sigma)$ are chosen so that

$$\int_a^b \tilde{f}(x) \psi_i(x) dx = 0 \quad (i = 1, 2, \dots, k).$$

Clearly, as $\sigma \rightarrow 0$,

$$A(\sigma) = \left\| \int_a^b \tilde{\psi}_i(x) \tilde{\psi}_j(x) dx \right\| \rightarrow A = \left\| \int_a^b \psi_i(x) \psi_j(x) dx \right\|$$

which is nonsingular, since $\psi_i(x)$ are linearly independent by assumption. Therefore, $a_i(\sigma)$ tend to zero ($i = 1, \dots, k$). Finally, we note that for all σ sufficiently small $\tilde{f}(x)$ exhibits exactly k sign changes or otherwise the hypothesis that f has only simple zeros would be contradicted. Now, appealing to Lemma A, we know that $\int_a^b \tilde{f}(x) \tilde{\varphi}(x) dx \geq 0$. Letting $\sigma \rightarrow 0$ we infer that

$$\int f(x) \varphi(x) dx \geq 0.$$

The discrete version of this approximation procedure applies in a simpler manner, described below.

LEMMA B'. Assume

$$\Psi \begin{pmatrix} x_1, \dots, x_j \\ 1, \dots, j \end{pmatrix} \geq 0$$

but not identically vanishing, $j = 1, 2, \dots, k$, and that $f(i)$ ($0 \leq i \leq N$) orthogonal to ψ_j has no more than k sign changes, in the sense that an arbitrary assignment of sign to any possible zeros in the range of f will produce no more than k strict sign changes. Let $f(N) \geq 0$, and if $f(N) = 0$, we assume the assignment of a positive value at N is compatible with achieving a maximum number of strict sign changes. In this case $\sum_0^N f(i) \varphi(i) \geq 0$ for all φ in the cone $C(\psi_1, \dots, \psi_k)$ defined by

$$\begin{vmatrix} \psi_1(x_1) & \dots & \psi_k(x_1) & \varphi(x_1) \\ \vdots & & \vdots & \vdots \\ \psi_1(x_{k+1}) & \dots & \psi_k(x_{k+1}) & \varphi(x_{k+1}) \end{vmatrix} \geq 0.$$

EXAMPLES.

1. Let a_j be defined by the generating function

$$(36) \quad \sum_{j=0}^n a_j x^j = \prod_{i=1}^n (q_i + p_i x) \quad p_i + q_i = 1; \quad 0 \leq p_i \leq 1.$$

Since $a_j / \binom{n}{j} p^j (1-p)^{n-j}$ is strictly logarithmically concave, i.e.,

$$\log a_j - \log \binom{n}{j} p^j (1-p)^{n-j}$$

is strictly concave in j unless constant [2, pp. 51, 52], we deduce that

$$(37) \quad b_j = a_j - \binom{n}{j} p^j (1-p)^{n-j} \quad j = 0, 1, \dots, n$$

changes sign at most twice. Note that this is true for every $0 < p < 1$. In particular if $np = p_1 + p_2 + \dots + p_n$ then the 0th and 1st moment of $(b_j)_{j=0}^n$ are zero and taking account of Lemma a, we conclude that $\{b_j\}$ changes sign exactly twice excluding the trivial case for $p_1 = \dots = p_n$. The conditions of Lemma b are thereby satisfied and again (8) is proved.

2. Consider

$$\psi_1(x) \equiv 1, \quad \psi_2(x) = \begin{cases} 0 & x = 0, 1, \dots, r \\ 1 & x = x + 1, \dots, n \end{cases} \quad r < n.$$

Now $\Psi \binom{x_1, x_2}{1, 2} \geq 0$ with inequality if $x_1 \leq r, x_2 > r$, permitting application of Lemma B'. The set of φ defined by (18) consists of those sequences of length n which are decreasing in the sub-interval $[0, r]$ and increasing in $[r + 1, n]$.

Let a_j be defined by (36) and determine \tilde{p} by the moment condition

$$\sum_{j=r+1}^n a_j = \sum_{j=r+1}^n \binom{n}{j} \tilde{p}^j (1 - \tilde{p})^{n-j}.$$

Then, apart from the trivial case $p_1 = \dots = p_n$, we know that the sequence (37) with p replaced by \tilde{p} changes sign twice (provided we adopt the convention that zero terms are given signs to produce the maximum number of sign changes).

It remains to examine the sign of the endpoint b_0 (or b_n). To accomplish this we consider two cases: (i) $(p_1 + \dots + p_n)/n = \bar{p} \leq \tilde{p}$ and so $\bar{p}^n \leq \tilde{p}^n$. But $a_n = p_1 p_2 \dots p_n < \bar{p}^n$ (recall that the case $p_1 = \dots = p_n$ is excluded) by the inequality between the geometric and arithmetic means so that $b_n = a_n - \tilde{p}^n < 0$. Similarly, if $\bar{p} > \tilde{p}$ then $\bar{q} < \tilde{q}$ and it follows analogously that $b_0 < 0$. This yields

$$\sum_{j=0}^n b(j) \varphi(j) \leq 0$$

for all φ in $C(\psi_1, \psi_2)$, i.e. for φ decreasing in $[0, r]$ and increasing in $[r + 1, n]$.

REMARK. While $a_j / \left[\binom{n}{j} p^j (1-p)^{n-j} \right]$ is logarithmically concave as a sequence in j for all p in $(0, 1)$, permitting the use of Lemma B', examples can be given showing that $\log(a_j/a'_j)$ is not necessarily concave where $\sum_0^n a'_j x^j = \prod_{i=1}^n (q_i + p_i x)$ and $(p') = (\pi)(p)$ for an arbitrary doubly stochastic matrix (π) . However, it follows from logarithmic concavity of Example 1 that whenever \tilde{p} can be adjusted so that

$$\sum_{j=0}^n \psi_2(j) \left(a_j - \binom{n}{j} \tilde{p}^j (1 - \tilde{p})^{n-j} \right) = 0$$

and $\psi_2(j)$ is non-constant and non-decreasing, then the above application of Lemma B' holds without change. As a consequence, we have

$$\sum_0^n \varphi(j) a_j \leq \sum_0^n \binom{n}{j} \tilde{p}^j (1 - \tilde{p})^{n-j} \varphi(j)$$

whenever φ is in $C(1, \psi_2)$.

We may combine the original Hoeffding result with the fact that $\sum_0^n \varphi(j) \binom{n}{j} p^j (1-p)^{n-j}$ is increasing in p whenever φ is increasing in j (cf. [5]) to observe that

$$\sum_{j=0}^n \varphi(j) a_j \leq \sum_{j=0}^n a(j) \binom{n}{j} \bar{p}^j (1 - \bar{p})^{n-j} < \sum_{j=0}^n \varphi(j) \binom{n}{j} p^j (1 - p)^{n-j}$$

whenever φ is both convex and increasing and $p > \bar{p} = (p_1 + \dots + p_n)/n$.

3. Let $\tilde{\lambda}$ be determined so that

$$\sum_{k=0}^r a_k = \sum_{k=0}^r \frac{\tilde{\lambda}^k}{k!} e^{-\tilde{\lambda}}.$$

Following the methods of the previous example, we obtain that

$$\sum_{k=0}^{\infty} a_k \varphi(k) \leq \sum_{k=0}^{\infty} \frac{\tilde{\lambda}^k}{k!} e^{-\tilde{\lambda}} \varphi(k)$$

for all sequences $\varphi(k)$ which are increasing on the segment $[0, r]$ and decreasing on $[r + 1, \infty)$.

There are continuous analogs of Examples 2 and 3.

REFERENCES

1. E. F. Beckenbach, and R. Bellman, *Inequalities*, Ergebnisse der Math. und Ihrer Grenzgebiete, Neue Folge, Heft 30, Springer-Verlag, 1961.
2. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, 2nd Ed., 1952.
3. ———, *Some simple inequalities satisfied by convex functions*, Mess. Math., **58** (1929), 145-152.

4. E. Hoeffding, *On the distribution of the number of successes in independent trials*, *Annals Math. Stat.*, **27**, No. 3 (1956), 713-721.
5. S. Karlin, *Total Positivity and Applications*, a forthcoming book.
6. ———, F. Proschan, and R. Barlow, *Moment inequalities of Pólya frequency functions*, *Pacific J. Math.*, **11**, No. 3 (1961), 1023-1033.
7. S. Karlin, *Total Positivity and Convexity Preserving Transformations*, *Proc. of Symposia in Pure Math. Vol. VII, Convexity*. Amer. Math. Soc., 1963.
8. ———, and H. Rubin, *The theory of decision procedures for distributions with monotone likelihood ratio*, *Ann. Math. Stat.*, **27** (1956), 272-299.
9. L. Mirsky, *On the trace of matrix products*, *Mathematische Nachrichten* (1959), 161-174.
10. G. Pólya, *On the mean-value theorem corresponding to a given linear homogeneous differential equation*, *Trans. Amer. Math. Soc.*, **24** (1922), 312-324.
11. I. J. Schoenberg, *On Pólya frequency functions*, I, *J. d'Analyse Math.*, I (1951), 331-374.

