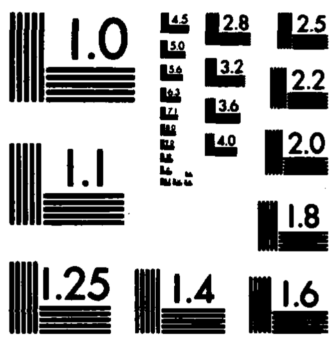
AD-A171 415 MIXTURES GENERALIZED CONVEXITY AND BALAYAGES(U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF STATISTICS J LYNCH JUN 86 TR-120 ARO-23448.1-MA DAAL03-86-K-0010

1/1

UNCLASSIFIED

F/G 12/1

NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

FOR REPRODUCTION PURPOSES

(2)

SECURIT

AD-A171 415

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ARO 23448.1-MA

2. GOVT ACCESSION NO.: N/A

3. RECIPIENT'S CATALOG NUMBER: N/A

4. TITLE (and Subtitle): Mixtures, Generalized Convexity and Balayages

5. TYPE OF REPORT & PERIOD COVERED: Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): James Lynch

8. CONTRACT OR GRANT NUMBER(s): DAAL03-86-K-0010

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
University of South Carolina
Columbia, SC 29208

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS:
U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

12. REPORT DATE: June, 1986

13. NUMBER OF PAGES: 20

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE AUG 2 7 1986 S D D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Mixtures, Tchebycheff system, Total positivity, Generalized convexity, Dilations, Balayages.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

How the dispersiveness of the mixing distribution carries over to the mixed model is qualified in terms of generalized convex functions. These ideas are extensions of those in Shaked (1980) and Schweder (1982). A representation akin to the one for dilations is also given for balayages defined in terms of these generalized convex functions.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ARO 23448.1-MA

Mixtures, Generalized Convexity and Balayages

by

James Lynch

University of South Carolina
Statistics Technical Report No. 120
60B05-14

June, 1986

Department of Statistics
University of South Carolina
Columbia, SC 29208

Research supported by the U. S. Army Research Office Grant No. DAAL03-86-K-0010.

Key Words: Mixtures, Tchebycheff system, Total positivity, Generalized convexity, Dilations, Balayages.

AMS 1980 Subject Classification Numbers: Primary: 60B05, 62E10
Secondary: 26A51, 46A55

Abstract

How the dispersiveness of the mixing distribution carries over to the mixed model is qualified in terms of generalized convex functions. These ideas are extensions of those in Shaked (1980) and Schweder (1982). A representation akin to the one for dilations is also given for balayages defined in terms of these generalized convex functions.

1. Introduction. In certain statistical problems, one typically has in mind a family $\{F_\theta: \theta \in \Theta\}$ of models (distributions) for the observations. As sometimes happens, though, the observed data may be "more dispersed" than might be expected of the above family. This could suggest that a "mixed model" may be a more appropriate fit since mixing introduces more dispersion intc the model.

In this paper we qualify just how "dispersiveness" in the mixing distribution carries over to the mixed model for certain types of models. This extends the work of Shaked (1980) and of Schweder (1982). More specifically (and ignoring obvious measure theoretic technicalities), for a mixing distribution λ on Θ, let $F_\lambda = \int F_\theta \, d\lambda$ denote the mixed model. When the models, F_θ, $\theta \in \Theta$, arise from a family of densities $\{f_\theta: \theta \in \Theta\}$ with respect to a σ-finite measure m, $f_\lambda = \int f_\theta \, d\lambda$ will denote the mixed density with respect to m. Note that $f_\theta \equiv f_{\delta_\theta}$ when δ_θ is the *mixing distribution degenerate* at θ.

Shaked (1980) investigated two types of dispersiveness for one parameter exponential families. One type was in terms of sign changes and the other in terms of dilations. (A distribution G is said to be a dilation of another distribution F, written $G \overset{d}{>} F$, if $\int c\,dF \le \int c\,dG$ for all convex c.) Shaked showed that $f_\lambda - f_{\theta^*}$ has two sign changes and the order is $+, -, +$ when λ satisfies the first "moment" condition $\int \tilde{u}(\theta) d\lambda(\theta) = \tilde{u}(\theta^*)$ where $\tilde{u}(\theta) = \int x f_\theta(x) dm(x)$. He also showed that if $\tilde{u}(\theta)$ is linear in θ and $\gamma \overset{d}{>} \lambda$, then $F_\gamma \overset{d}{>} F_\lambda$.

Schweder (1982) further investigated this second type of dispersiveness and showed that $F_\gamma \overset{d}{>} F_\lambda$ whenever $\gamma \overset{d}{>} \lambda$ if and only if the family $\{F_\theta: \theta \in \Theta\}$ is convexly parameterized. That is, $\tilde{c}(\theta) = \int c(x)dF_\theta(x)$ is convex whenever c is convex.

The above two types of dispersiveness might be considered first order notions of dispersiveness. The sign change since f_λ is compared to f_{θ^*} which arises from the degenerating mixing distribution δ_{θ^*}; the dilation since $\gamma \overset{d}{>} \lambda$ if and only if $\gamma(\cdot) = \int P(\cdot|\theta)d\lambda(\theta)$, where $P(.|\theta) \overset{d}{>} \delta_\theta$ is a probability distribution for each θ. (See Strassen, 1965, Theorems 2 and 8.)

Here we are interested in higher order (k-order, $k \geq 1$) notions of dispersiveness. These higher order notions involve Tchebycheff systems (T-systems) of functions $U = \{u, \ldots, u_{2k-1}\}$ and U-convex functions which are defined in terms of U.

In Section 2, a rudementary account on T-systems and U-convexity is given and a simple characterization of U-convexity is proved (Theorem 2.1). Very thorough accounts on T-systems and generalized convexity can be found in Karlin and Studden (1966) and in Karlin (1968). A palatable introduction to generalized convexity can be found in Roberts and Varberg (1973).

In Section 3, U-$\tilde{U}$ convexly parameterized families are defined for T-systems U and $\tilde{U}$. It is shown that $\{F_\theta: \theta \in \Theta\}$ is U-$\tilde{U}$ convexly parameterized if and only if $F_\nu \overset{U}{>} F_\lambda$ whenever $\gamma \overset{\tilde{U}}{>} \lambda$ where $\overset{U}{>}$ and $\overset{\tilde{U}}{>}$ are partial orderings defined in terms of U and $\tilde{U}$ (Theorem 3.1). In addition it is shown that under the (equivalent) moment conditions

$$\int \tilde{u}_j d\lambda = \int \tilde{u}_j d\lambda_k \qquad j=0,\ldots,2k-1$$

or

$$\int u_j dF_\lambda = \int u_j dF_{\lambda_k} \qquad j=0,\ldots,2k-1$$

$f_\lambda - f_{\lambda_k}$ has 2k sign changes and the order is $+,-,+,\ldots,-,+$ where λ is discrete with k mass points (Theorems 3.2, 3.3 and 3.4). The latter result is useful in determining "if you've gone for enough" when fitting a mixed model using a method of moments approach.

Finally, in Section 4, a necessary and sufficient condition is given to show when a probability measure γ has the representation

$$\gamma(\cdot) = \int P(\cdot|x_1,\ldots,x_k) \prod_1^k d\lambda(x_i)$$

where $$P(\cdot|x_1,\ldots,x_k) \overset{U}{>} F_{\underline{x}}$$

and $F_{\underline{x}}$ is the empirical distribution function for the sample $\underline{x} = (x_1,\ldots,x_k)$ (Theorem 4.1).

2. U-Convexity. Fundamental to the notion of U-convexity is the definition of a Tchebycheff system. (Throughout this section, $X = \{x_i: i=0,1,\ldots,n+1\}$, $x_0 < x < \ldots < x_{n+1}$.)

Definition. A family of functions $U = \{u_i: i=0,1,\ldots,n\}$ defined on X is said to be a Tchebycheff system (T-system) on X if the determinant

$$u(X') \equiv u(x_0',\ldots,x_n') \equiv \begin{vmatrix} u_0(x_0') & \cdots & u_0(x_n') \\ u_1(x_0') & \cdots & u_1(x_n') \\ & \vdots & \\ u_n(x_0') & \cdots & u_n(x_n') \end{vmatrix}$$

is positive whenever $X' = \{x_0' < \ldots < x_n'\} \subset X$. For a set Y of cardinality greater than $n+1$, the family U is said to be a T-system on Y if U is a T-system for each $X \subset Y$.

Definition. Let $U = \{u_i: i=0,\ldots,n\}$ be a T-system on X. A function f is said to be U-convex on X if the determinant

$$u_f(X) \equiv \begin{vmatrix} u_0(x_0) & \cdots & u_0(x_{n+1}) \\ u_1(x_0) & \cdots & u_1(x_{n+1}) \\ & \vdots & \\ u_n(x_o) & \cdots & u_n(x_{n+1}) \\ f(x_0) & \cdots & f(x_{n+1}) \end{vmatrix} \geq 0.$$

If U is a T-system on a set Y of cardinality greater than $n+1$, f is said to be U-convex on Y if f is U-convex on each $X \subset Y$. A function f is said to be U-concave if $-f$ is U-convex.

Remark. Note that a polynomial in the u's, $P(x) = A_0u_0(x) + A_1u_1(x) + \ldots + A_nu_n(x)$, is both U-convex and U-concave.

The next theorem gives a useful characterization of U-convexity. For the usual definition of convexity, i.e., $u_0 \equiv 1$ and $u_1(x) \equiv x$, it corresponds to the midpoint of the chord between two points on the graph of a convex function lying above the function.

For this characterization we need the following notation. Let $k = \left[\frac{n+2}{2}\right]$ where $[x]$ denotes the integer part of x. For $X = \{x_0 < x < \ldots < x_{n+1}\}$, let $t_k = x_n$, $t_{k-1} = x_{n-2}$, $t_{k-2} = x_{n-4}, \ldots$, i.e., $t_{k-j} = x_{n-2j}$ for $j = 0,1,\ldots,k-1$. For $\underline{t} = (t_1,t_2,\ldots,t_k)$, let $F_{\underline{t}}$ denote both the probability distribution and probability measure which places mass k^{-1} at t_i. $F_{\underline{t}}$ is just the empirical distribution for the sample $t_1,\ldots,t_k$.

<u>Theorem 2.1</u>. A function f is U-convex on X if and only if

(2.1) $$\int f dF_{\underline{t}} \le \int f d\lambda$$

for each finite measure λ with support contained in X satisfying

(2.2) $$\int u_j dF_{\underline{t}} = \int u_j d\lambda \quad \text{for } j=0,1,\ldots,n.$$

<u>Proof</u>. ($\rightarrow$) If f is not U-convex, then $u_f(X) < 0$. So,

(2.3) $$A\underline{\Delta} \equiv \begin{bmatrix} u_0(x_0) & \cdots & u_0(x_{n+1}) \\ u_1(x_0) & \cdots & u_1(x_{n+1}) \\ & \vdots & \\ u_n(x_0) & \cdots & u_n(x_{n+1}) \\ f_0(x_0) & \cdots & f(x_{n+1}) \end{bmatrix} \begin{bmatrix} \Delta_0 \\ \Delta_1 \\ \vdots \\ \Delta_n \\ \Delta_{n+1} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix} \equiv -1\,\underline{e}$$

has a solution $\underline{\Delta}$.

By Cramer's rule, $\Delta_j = (-1)^{n+4+j}\, u(X_j) / u_f(X)$ where $X_j = X - \{x_j\}$. Since $u_f(X) < 0 < u(X_j)$, Δ_j alternates in sign with $\Delta_{n+1} < 0$. So,

$c = \max\{\Delta_j: j=n,n-2,\dots\} > 0.$

Let $\alpha_j = c - \Delta_j$ for $j=n,n-2,\dots$ and $= -\Delta_j$ for the other values of j. Then, by (2.3),

$$0 = \sum_{j=0}^{n+1} u_i(x_j)\Delta_j = c \sum_{j=n,n-2,\dots} u_i(x_j) - \sum_{j=0}^{n+1} u_i(x_j)\alpha_j$$

and

$$-1 = \sum_{j=0}^{n+1} f(x_j)\Delta_j = c \sum_{j=n,n-2,\dots} f(x_j) - \sum_{j=0}^{n+1} f(x_j)\alpha_j.$$

Setting $\lambda(\{x_j\}) = \alpha_j(kc)^{-1} \geq 0$, we have from the above that (2.2) is satisfied but $\int f d\lambda < \int f F_{\underline{t}}$. This proves the "if" part of the theorem.

($\leftarrow$) Now let f be U-convex. If $u_f(X) = 0$, then f is a polynomial in the u's. In this case, from (2.2) equality holds in (2.1). Thus to complete the proof, we only need to consider when $u_f(X) > 0$.

Let λ denote a measure whose support is contained in X and which satisfies (2.2). Let $\Delta_i = \Delta(\{x_i\}) = \lambda(\{x_i\}) - F_{\underline{t}}(\{x_i\})$ and $c = \int f d\Delta$. Then for A and $\underline{e}$ as defined in (2.3), $A\underline{\Delta} = c\underline{e}$. So, from Cramer's rule, $0 \leq \lambda(\{x_{n+1}\}) = \Delta_{n+1} = c(-1)^{2(n+2)}u(X_{n+1}) / u_f(X)$. Since $u_f(X) > 0$ and $u(X_{n+1}) > 0$, it follows that $\int f d\lambda - \int f dF_{\underline{t}} = c \geq 0$. ∎

U-convex functions can be used to define a measure of dispersiveness for probability measures. This is needed in the next section to qualify how dispersiveness of the mixing distribution carries over to the mixed model. The terminology is from Meyer (1966).

<u>Definition</u>. Let $U = \{u_0,\dots,u_n\}$ be a T-system on a Borel set $Y \subset R$. Let λ and ν be two finite measures on Y. If $\int f d\lambda \leq \int f d\nu$ for all integrable U-convex f, then ν is called a <u>balayage</u> of λ. This is written as $\lambda \overset{U}{<} \nu$ or $\nu \overset{U}{>} \lambda$. Note that if $u=1$ is in U, then $\int d\lambda = \int d\nu$.

3. U-$\tilde{U}$ Convexly Parameterized Families. Let $\{F_\theta: \theta \in \Theta\}$ be a family of distribution functions on $X \subset R$ where $\Theta \subset R$. For a (integrable) function g, let $\tilde{g}(\theta) = \int g(x)dF_\theta(x)$.

Definition. Let $U = \{u_0,\ldots,u_n\}$ be a T-system on X and let $\tilde{U} = \{\tilde{u}_0,\tilde{u}_1,\ldots,\tilde{u}_n\}$. The family $\{F_\theta: \theta \in \Theta\}$ is said to be U-$\tilde{U}$ convexly parameterized if (1) $\tilde{U}$ is a T-system on Θ, and (2) $\tilde{c}$ is $\tilde{U}$-convex whenever c is U-convex. (Implicit here is that $u_j(x)$ is integrable for each F_θ and that the cardinalities of X and of Θ are greater than n.)

Example 1. Let F_θ be absolutely continuous with respect to some σ-finite measure m on X. Let $f_\theta = dF_\theta|dm$. If $f_\theta(x)$ is strictly totally positive (STP) of order n+1, (see Karlin, 1968, pages 11 and 12 for the definition), then $\tilde{U}$ is a T-system whenever U is a T-system. This follows from the basic composition formula on page 98 of Karlin (1968) (see also Theorem 3.2 on page 284).

Example 2. The one parameter exponential family with density $f_\theta(x) = e^{x\theta}B(\theta)$ is STP of all orders up to the minimum of the cardinalities of Θ and X. Such a family includes the binomial family, the Poisson, the gamma with fixed shape parameter, and the normal with fixed variance. See Karlin (1968), page 19, for details.

Analogous to Schweder's (1982) theorem on on page 166 for convexly parameterized families, the following theorem points out the connection between U-$\tilde{U}$ convexly parameterized families and balayages.

Theorem 3.1. Let $U = \{u_0,\ldots,u_n\}$ be a T-system for which $\tilde{U}$ is a T-system for the family $\{F_\theta: \theta \in \Theta\}$. Then $\{F_\theta: \theta \in \Theta\}$ is $U - \tilde{U}$ convexly parameterized if and only if $F_\lambda \overset{U}{<} F_\nu$ whenever $\lambda \overset{\tilde{U}}{<} \nu$.

<u>**Proof**</u>. ($\rightarrow$) Let $\Theta_{n+1} = \{\theta_0 < \theta_1 < \ldots < \theta_{n+1}\} \subset \Theta$. For $k = [\frac{n+2}{2}]$ and $j=0,1,\ldots,k-1$, let $t_{k-j} = x_{n-2j}$. Let $F_{\underline{t}}$ denote the probability distribution placing mass $1/k$ at each of the points $t_1,\ldots,t_k$ and let λ be any other finite measure with support contained in Θ_n and satisfying

$$\int \tilde{u}_j d\lambda = \int \tilde{u}_j dF_{\underline{t}} \qquad j=0,\ldots,n.$$

Then, by Theorem 2.1, $F_{\underline{t}} \overset{\tilde{U}}{<} \lambda$. So, $F_{F_{\underline{t}}} \overset{U}{<} F_\lambda$. Thus, if c is U-convex,

$$\int \tilde{c} dF_{\underline{t}} = \iint c(x) dF_\theta(x) dF_{\underline{t}}(\theta)$$

$$= \int c(x) dF_{F_{\underline{t}}}(x) \le \int c(x) dF_\lambda(x)$$

$$= \iint c(x) dF_\theta(x) d\lambda(\theta) = \int \tilde{c} d\lambda$$

This with another application of Theorem 2.1 yields that $\tilde{c}$ is $\tilde{U}$-convex.

($\leftarrow$) Let $\lambda \overset{\tilde{U}}{<} \nu$ and let c be U-convex. Since $\{F_\theta: \theta \in \Theta\}$ is U-$\tilde{U}$ convexly parameterized, $\tilde{c}$ is $\tilde{U}$ convex. So,

$$\int c dF_\lambda = \iint c(x) dF_\theta(x) d\lambda(\theta)$$

$$= \int \tilde{c} d\lambda \le \int \tilde{c} d\nu$$

$$= \iint c(x) dF_\theta(x) d\nu(\theta) = \int c dF_\nu.$$

Consequently, $F_\lambda \overset{U}{<} F_\nu$. ∎

In the next three theorems sign change results are given for $f_\lambda - f_{\lambda_k}$ when

(3.1) $$\int \tilde{u}_j d\lambda = \int \tilde{u}_j \, d\lambda_k \quad \text{for } j=0,1,\ldots,2k-1,$$

and λ_k is discrete with k mass points. In these three theorems it is assumed that, for each $\theta \in \Theta$, F_θ has a density f_θ with respect to a σ-finite measure m which is STP_{2k+1} on $\Theta \times X$, X the support of m. Throughout it also is assumed that, for each j, u_j is integrable with respect to f_λ and f_{λ_k}.

The first theorem deals with the classical T-system $U = \{1,x^1,\ldots,x^{2k-1}\}$ and generalizes Theorem 1 of Shaked (198).

Theorem 3.2. Let λ and λ_k be two mixing distributions satisfying (3.1) for $U = \{1,x^1,\ldots,x^{2k-1}\}$ where λ_k is discrete with k mass points. If $m(\{f_\lambda \neq f_{\lambda_k}\}) > 0$, then $f_\lambda - f_{\lambda_k}$ has 2k sign changes on X and the order is $+,-,+,\ldots,-,+$.

Proof. Note that from the definition of STP_{2k+1} it is implicit in the statement of the theorem that both Θ and X are of cardinality greater than 2k.

For θ a mass point of λ_k, let $s(\theta) = -1$ if $\lambda(\{\theta\}) \leq \lambda_k(\{\theta\})$ and let $s(\theta) = 1$ otherwise. So $s(\theta)$ has at most 2k sign changes.

Let μ be the measure given by $d\mu = s(\cdot)d(\lambda-\lambda_k)$. Since $\Delta(x) \equiv f_\lambda(x) - f_{\lambda_k}(x) = \int s(\theta)f_\theta(x)d\mu(\theta)$ and $f_\theta(x)$ is STP_{2k+1}, it follows from the variation diminishing theorem (Karlin, 1968, page 233) that $\Delta(\cdot)$ can have at most 2k sign changes. If there are less than 2k sign changes, say l sign changes, then there are l points in X, $x_1 < x_2 < \ldots < x_l$, such that $\Delta(x)\Delta(y) \leq 0$ when $x \in I_j$ and $y \in I_{j+1}$, $j=0,\ldots,l-1$ and $I_0 = (-\infty,x_1)$, $I_1 = (x_1,x_2)$, ..., $I_l = (x_l,\infty)$. Let $P(x) = (x-x_1)(x-x_2)\ldots(x-x_l)$. Since $P(x)$ is a polynomial of degree $l \leq 2k-1$, it follows from (3.1) that

$$\int P(x)\Delta(x)dm(x) = 0. \tag{3.2}$$

Since $P(x)\Delta(x)$ is of the same sign and $P(x) \neq 0$ except at $x_1,\ldots,x_l$, (3.3) $\Delta(x) = 0$ a.e. $[m]$ on $X - X_0$, $X_0 = \{x_1, \ldots, x_l\}$ from (3.2).

Thus, for $m_n = \Delta(x_n) m(\{x_n\})$,

$$0 = \sum_{n=1}^{l} x_n^j \, m_n, \quad j=0,\ldots,l-1$$

from (3.1) and (3.3). So $m \equiv 0$. This with (3.3) implies that $f_{\lambda_k} = f_\lambda$ a.e. $[m]$ which contradicts the hypotheses of the theorem. ∎

When $U = \{u_0, \ldots, u_{2k-1}\}$ is a <u>Haar system</u>, i.e. $\{u_0, \ldots, u_j\}$ is a T-system for $j=0,1,\ldots,2k-1$, then the next theorem is a consequence of Theorem 5.2 on page 30 of Karlin and Studden (1966) and the above proof with x^j replaced by $u_j(x)$.

<u>Theorem 3.3</u>. Assume that the support of m, X, is contained in a finite interval $[a,b]$. Let $U = \{u_0, u_1, \ldots, u_{2k-1}\}$ be a Haar system, of continuous functions on $[a,b]$. Let λ and λ_k be two mixing distributions satisfying (3.1) where λ_k is discrete with k mass points. If $m(\{f_\lambda \neq f_{\lambda_k}\}) > 0$, then $f_\lambda - f_{\lambda_k}$ has $2k$ sign changes on X and the order is $+,-,+,\ldots,-,+$.

For the next theorem, it is assumed that $U = \{u_0, u_1, \ldots, u_{2k}\}$ is a <u>Descartes system</u>, i.e., $\{u_{i_1}, u_{i_2}, \ldots, u_{i_m}\}$ is a T-system for each $\{i_1, < \ldots <, i_m\} \subset \{0, \ldots, 2k\}$.

<u>Theorem 3.4</u>. Let $U = \{u_0, \ldots, u_{2k}\}$ be a Descartes system on X. Let λ and λ_k be two mixing distributions satisfying (3.1) where λ_k is discrete with k mass points. If $m(\{f_\lambda \neq f_{\lambda_k}\}) > 0$, then $f_\lambda - f_{\lambda_k}$ has $2k$ sign changes on X and the order is $+,-,+,\ldots,-,+$.

<u>Proof</u>. As in the first part of the proof of Theorem 3.2, $\Delta \equiv f_\lambda - f_{\lambda_k}$ has at most $2k$ sign changes by the variation diminishing theorem (page 233 of Karlin, 1968).

Since U is a Descartes system, $u_j(x)$ is STP_{2k+1} on $\{0,1,\ldots,2k\} \times X$. If Δ has less than $2k$ sign changes, say $l \leq 2k-1$ sign changes, another application of the variation diminishing theorem shows that $g(j) = \int u_j(x)\Delta(x)dm(x)$ can have at most l sign changes on $\{0,1,\ldots,2k\}$ where zeroes of g can be arbitrarily assigned either sign. But this leads to a contradiction since $g(j) = 0$ for $j=0,\ldots,2k-1$. ∎

Remark. These Theorems should be compared with Theorems 5.4 and 5.5 on pages 409 and 410 of Karlin and Studden (1966). Note that there U is an extended complete T-system (or what might be called an extended Haar system) which involves assumptions on the derivatives of the u's.

<u>4. A Representation Theorem</u>. For k a fixed positive integer, let $U = \{u_0, u_1, \ldots, u_{2k-1}\}$ be a T-system of continuous functions on $I=(a,b)$ an open interval. When $k > 1$ it shall be further required that U be an <u>extended T-system</u>, i.e., in addition to U being a T-system, each $u_j \in C^{2k-1}(I)$ and, for l distinct values of the x's $(l = 1,\ldots,2k-1)$,

$$a<x_0 = x_1 = \ldots = x_{q_1} < x_{q_1+1} = x_{q_1+2} \ldots = x_{q_2} <\ldots< x_{q_{l-1}+1} = \ldots = x_{q_l} < b, \; q_l$$

$= 2k-1$, the following determinants are all positive:

$$u^*(x_0, x_1, \ldots, x_{2k-1}) =$$

$$\begin{vmatrix} u_0(x_{q_1}) & u_0'(x_{q_1}) & \ldots u_0^{(q_1)}(x_{q_1}) & u_0(x_{q_2}) & \ldots u_0(x_{q_l}) & \ldots u_0^{(q_l-q_{l-1}+1)}(x_{q_l}) \\ u_1(x_{q_1}) & u_1'(x_{q_1}) & \ldots u_1^{(q_1)}(x_{q_1}) & u_1(x_{q_2}) & \ldots u_1(x_{q_l}) & \ldots u_1^{(q_l-q_{l-1}+1)}(x_{q_l}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{2k-1}(x_{q_1}) & u_{2k-1}'(x_{q_1}) & \ldots u_{2k-1}^{(q_1)}(x_{q_1}) & u_{2k-1}(x_{q_2}) & \ldots u_{2k-1}(x_{q_l}) & \ldots u_{2k-1}^{(q_l-q_{l-1}+1)}(x_{q_l}) \end{vmatrix}$$

(See Karlin and Studden, 1966, page 6.) In this section a representation is obtained for balayages defined in terms of U-convex functions which is akin to the (Hardy-Littlewood-Polya-Blackwell-Stein-Sherman-Cartier-Fell-Meyer-Strassen) representation for dilations (see Strassen, 1965, Theorems 2 and 8).

To state the representation theorem requires the following notation. Let $F = \{f\colon f$ is U-concave on $I\}$. Note that since U is a T-system of continuous functions, any $f \in F$ is continuous. Furthermore, when U is an ET-system with $k > 1$, $f \in F$ is differentiable. (Theorems B and D on pages 248 and 249 of Roberts and Varberg, 1973, or Theorem 3.4 on page 188 of Karlin, 1968).

For $\underline{x} \in I^k$ and f a real valued function on I, let $\bar{f}(\underline{x}) = (f(x_1) + \ldots + f(x_k))/k$. Let B and B^k denote the Borel subsets of R and R^k, respectively. For ν a probability measure (p.m.) on (R,B), let $S(\nu) \equiv \{x: \nu((x-\varepsilon,x+\varepsilon)) > 0$ for every $\varepsilon > 0\}$ denote the support of ν. Note that $S(\nu)$ is always closed.

The following conditions are imposed in the representation theorem:

(c1) ν and λ are two p.m.'s on (R,B) with supports contained in a compact interval $K \subset I$ and satisfying

(c2) $\int f_1 \wedge \ldots \wedge f_m d\nu \leq \int \bar{f}_1 \wedge \ldots \wedge \bar{f}_m \Pi_1^k \, d\lambda$

whenever $f_i \in F$, $i=1,\ldots,m, m=1,2,\ldots$.

Below $P(\cdot|\cdot)$ denotes a <u>Markov kernel</u> on $B \times K^k$, i.e., $P(\cdot|\cdot)$ is a p.m. on (R,B) for each $\underline{x} \in K^k$ and $P(A|\cdot)$ is B_{K^k} ($\equiv$ {Borel subsets of K^k}) measurable for each $A \in B$.

<u>Theorem 4.1</u>. Under condition (c1), (c2) is necessary and sufficient for $\nu(A) = \int P(A|\underline{x}) \, \Pi_1^k \, d\lambda(x_i)$ for every $A \in B$ where $P(\cdot|\cdot)$ is a Markov kernel on $B \times K^k$ with $P(\cdot|\underline{x}) \overset{U}{>} F_{\underline{x}}$ for every $\underline{x} \in K^k$.

The proof of Theorem 4.1, though somewhat involved, is really along the line of Strassen's (1965) proof for dilations. Before giving the proof some further quantities need to be defined and some lemmas need to be stated and proved.

Let K be a compact interval contained in I. Later K will be chosen to contain $S(\nu)$ and $S(\lambda)$. Let D denote the set of discrete p.m.'s on (K,B_K) with at most k mass points.

Let M denote the moment space $\{\underline{m} \in R^{2k}: m_j = \int u_{j-1} dD,\ j=1,\ldots,2k, D \in D\}$.

Note that Theorem 2.1 and case 2 (ii) on pages 42 and 46, respectively, of Karlin and Studden (1966) guarantee that if $\underline{m} \in M$ are the "moments" of a p.m. with support contained in K, then there is a (unique) $D_{\underline{m}} \in D$ with moments $\underline{m}$. Consequently M is convex and, since the u's are continuous, it is easy to see that M is compact.

Let $f \in C(K)$. For $\underline{m} \in M$ and $\underline{x} \in K^k$, let

$$l_f(\underline{m}) = \sup\{\int f d\mu \colon \mu \overset{U}{>} D_{\underline{m}},\ S(\lambda) \subset K\}$$

and

$$h_f(\underline{x}) = \sup\{\int f d\mu \colon \mu \overset{U}{>} F_{\underline{x}},\ S(\lambda) \subset K\}$$

where $F_{\underline{x}}$ is the empirical distribution of the sample $x_1, x_2, \ldots, x_k$. Note that in the definition of l_f and h_f, μ is a p.m. since $u_0 \equiv 1$.

Let $m(\cdot)\colon K^k \to M$ be given by $m_j = \int u_{j-1} dF_{\underline{x}} = \sum_{i=1}^{k} u_{j-1}(x_i)/k$. Obviously $m(\cdot)$ is continuous and it follows from the definition of l_f and h_f that $h_f(\underline{x}) = l_f(m(\underline{x}))$.

In Lemma 4.2, the relative interior of M refers to the interior of M when M is viewed as a subset of the smallest affine set containing it (see Rockafellar, 1970).

<u>Lemma 4.2</u>. $l_f(\cdot)$ is concave on M, and consequently, continuous on the relative interior of M.

<u>Proof</u>. That $l_f(\cdot)$ is continuous on the relative interior of M is immediate from Theorem 10.1 of Rockafellar (1970) once $l_f(\cdot)$ is shown to be concave on M.

To do this, let $\underline{m}_1$ and $\underline{m}_2 \in M$ ($m_1 \neq m_2$), $\alpha \in (0,1)$ and $\bar{\alpha} = 1-\alpha$. Since M is convex, $\underline{m}_3 = \alpha\,\underline{m}_1 + \bar{\alpha}\,\underline{m}_2 \in M$. For i=1 and 2, let $\lambda_i \overset{U}{>} D_{\underline{m}_i}$ and let

$\lambda_3 = \alpha\,\lambda_1 + \bar{\alpha}\,\lambda_2$. That l_f is concave on M will follow immediately from the definition of l_f once it is shown that $\lambda_3 \overset{U}{>} D_{\underline{m}_3}$. Since $\overset{U}{>}$ is transitive, it suffices to show that $D \equiv \alpha\,D_{\underline{m}_1} + \bar{\alpha}\,D_{\underline{m}_2} \overset{U}{>} D_{\underline{m}_3}$.

Case 1: k=1. Let $x < y < z$ denote the three mass points of D and $D_{\underline{m}_3}$ and let g be U-convex. To avoid trivialities, assume that $u_g(x,y,z) > 0$.

First we show that y is the mass point corresponding to $D_{\underline{m}_3}$. If not assume that x is the mass point corresponding to $D_{\underline{m}_3}$. Then

$$\begin{bmatrix} 1 & 1 & 1 \\ u_1(x) & u_1(y) & u_1(z) \\ g(x) & g(y) & g(z) \end{bmatrix} \begin{bmatrix} -1 \\ \alpha \\ \bar{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix}, \quad c = \int g\,d(D - D_{\underline{m}_3}).$$

By Cramer's rule, $-1 = c\times u(y,z)/u_g(x,y,z)$ and $0 < \bar{\alpha} = c\times u(x,y)/u_g(x,y,z)$ which is a contradiction since $u(x,y)$, $u(y,z)$ and $u_g(x,y,z)$ are all positive. Similarly z cannot be a mass point of $D_{\underline{m}_3}$.

Since y is the mass point corresponding to $D_{\underline{m}_3}$,

$$\begin{bmatrix} 1 & 1 & 1 \\ u_1(x) & u_1(y) & u_1(z) \\ g(x) & g(y) & g(z) \end{bmatrix} \begin{bmatrix} \alpha \\ -1 \\ \bar{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix}, \quad c = \int g\,d(D - D_{\underline{m}_3}).$$

Again by Cramer's rule, $0 < \alpha = c\times u(y,z)/u_g(x,y,z)$. So, $0 < c$ since $u(y,z)$ and $u_g(x,y,z)$ are both positive.

Case 2: k > 1. Let $x_i \in K$, $i=1,\ldots,l \le k$ denote the mass points of $D_{\underline{m}_3}$. If $l < k$, let $x_i \in K$, $i=k-l+1,\ldots,k$ be chosen so that $x_1,\ldots,x_k$ are all distinct. Let $y_1 < y_2 < \ldots < y_k$ denote the ordered x's.

Let g be U-convex. Since U is an ET-system, recall that g is differentiable and $u^*(y_1,y_1,y_2,y_2,\ldots,y_k,y_k) > 0$. So there exists a polynomial

$P(x)$ in the u's such that $P(x_i) = g(x_i)$ and $P'(x_i) = g'(x_i)$ for i-1,...,k. By Theorem 2.2 on pages 282 and 283 of Karlin (1968), $g(x) \geq P(x)$ on K. Thus, since the "moments" of D agree with those of $D_{\underline{m}_3}$,

$$\int g dD \geq \int P dD = \int P dD_{\underline{m}_3} = \int g dD_{\underline{m}_3},$$

where the last equality follows since $P = g$ on the $S(D_{\underline{m}_3})$. So, $D \overset{U}{>} D_{\underline{m}_3}$. ▓

Lemma 4.3. $h_f(\underline{x})$ is continuous on K^k with $h_f \geq \bar{f}$.

Proof. That $h_f \geq \bar{f}$ is immediate from the definition of h_f.

Let $\underline{x} \in K^k$. If the coordinates of $\underline{x}$ are all distinct, it follows from Theorem 2.1 on page 42 of Karlin and Studden (1966) that $m(\underline{x})$ must be in the relatively interior of M. Since $h_f(\underline{x}) = l_f(m(\underline{x}))$, it follows from Lemma 4.2 that h_f is continuous at $\underline{x}$.

Now consider the case when at least two coordinates of $\underline{x}$ agree. Let $y_1 < \ldots < y_l$, $l < k$ denote the distinct values of $x_1, \ldots, x_k$. First we show that $\lambda = F_{\underline{x}}$ if $\lambda \overset{U}{>} F_{\underline{x}}$ with $S(\lambda) \subset K$, in which case, $h_f(\underline{x}) = \bar{f}(\underline{x})$.

Since $l < k$, by Theorem 5.2 on page 30 of Karlin and Studden there exists a nonnegative polynomial $P(x)$ in the u's whose only zeroes on K are $y_1, \ldots, y_l$. Since $P(\cdot)$ is both U-concave and U-convex, $\int P d(\lambda - F_{\underline{x}}) = 0$ whenever $\lambda \overset{U}{>} F_{\underline{x}}$ and $S(\lambda) \subset K$. But this implies that $S(\lambda) \subset \{y_1, \ldots, y_l\}$. So $\lambda \in D$. Since the "moments", $\underline{m}$, uniquely determine $D_{\underline{m}}$ and $\lambda \overset{U}{>} F_{\underline{x}}$ implies that $\int u_j d\lambda = \int u_j dF_{\underline{x}}$ for each j, $\lambda = F_{\underline{x}}$.

Let $\underline{x}_n \to \underline{x}$ through points in K^k as $n \to \infty$. Now we will show that $h_f(\underline{x}_n) \to \bar{f}(\underline{x}) = h_f(\underline{x})$. Let $\varepsilon_n \downarrow 0$, $\varepsilon_n > 0$ as $n \to \infty$. Choose $\mu_n \overset{U}{>} F_y$ with $S(\mu_n) \subset K$ such that $h_f(\underline{x}_n) \leq \int f d\mu_n + \varepsilon_n$. Since $f \in C(K)$, it suffices to show that $\mu_n \to F_{\underline{x}}$ in distribution since then

$$\bar{f}(\underline{x}) = \lim \bar{f}(\underline{x}_n) \leq \underline{\lim}\, h_f(\underline{x}_n)$$
$$\leq \overline{\lim}\, h_f(\underline{x}_n) \leq \lim \int f d\mu_n = \int f dF_{\underline{x}} = \bar{f}(\underline{x}).$$

Since K is compact, $\{\mu_n\}$ is tight. Thus to show that $\mu_n \to F_{\underline{x}}$ in distribution it suffices to show that if $\{\mu_m\}$ is a convergent subsequence of $\{\mu_n\}$, say converging to μ, then $\mu \overset{U}{>} F_{\underline{x}}$.

To see this, let g be U-convex. Since g is U-convex, g is continuous on K. Since $\mu_m \overset{U}{>} F_{\underline{x}_m}$,

$$\int g d\mu = \lim \int g d\mu_m \geq \lim \int g dF_{\underline{x}_m} = \int g dF_{\underline{x}}.$$

So $\mu \overset{U}{>} F_{\underline{x}}$. ∎

Now let $\mathbb{K}$ denote the set of functions on K^k which are uniform limits on K^k of functions of the form $\bar{f}_1 \wedge \bar{f}_2 \wedge \ldots \wedge \bar{f}_m$, $f_i \in F$, $i=1,\ldots,m$, $m=1,2,\ldots$. For two p.m.'s $\underline{\lambda}$ and $\underline{\nu}$ on (K^k, B_{K^k}) write $\underline{\nu} \overset{*}{>} \underline{\lambda}$ if $\int \underline{f} d\underline{\nu} \leq \int \underline{f} d\underline{\lambda}$ for every $\underline{f} \in \mathbb{K}$, i.e., $\underline{\nu}$ is a balayage of $\underline{\lambda}$ under $\overset{*}{>}$. Let $\delta_{\underline{x}}$ denote the p.m. which is degenerate at $\underline{x}$.

The following lemma characterizes $\mathbb{K}$ in terms of balayages of $\delta_{\underline{x}}$. The proof is the same as the proof of Theorem 47 on page 240 of Meyer (1966). As a corollary, we get that $h_f \in \mathbb{K}$.

<u>Lemma 4.4</u>. Let $\underline{f} \in C(K^k)$. Then $\underline{f} \in \mathbb{K}$ if and only if $\int \underline{f} d\underline{\lambda} \leq f(\underline{x})$ whenever $\underline{\lambda} \overset{*}{>} \delta_{\underline{x}}$, $\underline{x} \in K^k$.

<u>Corollary 4.5</u>. $h_f \in \mathbb{K}$.

<u>Proof</u>. Note that $h_f \in C(K^k)$ by Lemma 4.3. So, by Lemma 4.4, it suffices to show that $\int h_f d\underline{\lambda} \leq h_f(\underline{x})$ whenever $\underline{\lambda} \overset{*}{>} \delta_{\underline{x}}$ and $\underline{x} \in K^k$.

Since K is compact, it is easy to see that $h_f(\underline{x})$ is a <u>support function</u>

on C(K) (i.e., subadditive and nonnegative homogeneous as a function in f) for each $\underline{x}$ satisfying the conditions of Strassen's (1965) Theorem 1 (Theorem 51 on page 244 of Meyer, 1966). Thus $h(f) = \int h_f d\underline{\lambda}$ is the support function of p.m.'s of the form

$$(4.1) \qquad \nu(A) = \int P(A|\underline{y})d\underline{\lambda}(\underline{y}) \text{ where } P(\cdot|\underline{y}) \overset{U}{>} F_{\underline{y}} \text{ is a Markov kernel on } B_K \times K^k.$$

If $g \in F$, then $\bar{g} \in K$ and it is immediate from (4.1) that, for such a ν, $\int g d\nu = \iint g(z)P(dz|\underline{y})d\underline{\lambda}(\underline{y}) \le \int \bar{g}(\underline{y})d\underline{\lambda}(\underline{y}) \le \bar{g}(\underline{x}) = \int g dF_{\underline{x}}$ whenever $\underline{\lambda} \overset{*}{>} \delta_{\underline{x}}$. So, $\nu \overset{U}{>} F_{\underline{x}}$. Thus, since $h(f) = \sup\{\int f d\nu\colon \nu$ is of the form (4.1)$\}$ (by the Hahn-Banach Theorem - see (5) on page 424 of Strassen, 1964), $h(f) \le h_f(\underline{x})$. ∎

For a p.m. ν on (R,B) let ν_0 be the p.m. on (R^k,B^k) defined by $\nu_0(B_1 \times \ldots \times B_k) = \nu(B_1 \cap B_2 \cap \ldots \cap B_k)$ for $B_i \in B$, $i=1,\ldots,k$. In other words, ν_0 is just the p.m. which is concentrated on the diagonal of R and having univariate marginals ν.

<u>Lemma 4.6</u>. Let ν and λ be two p.m.'s on (R,B) with $S(\nu) \cup S(\lambda) \subset K$. Then condition (c2) is equivalent to

$$(c2') \quad \int \underline{f} d\nu_0 \le \int \underline{f}\ \Pi_1^k\, d\lambda \text{ for every } \underline{f} \in K.$$

<u>Proof</u>. From the definition of ν_0 it is clear that

$$(4.2) \quad \int \bar{f}_1 \wedge \ldots \wedge \bar{f}_m\, d\nu_0' = \int f_1 \wedge \ldots \wedge f_m\, d\nu \le \int \bar{f}_1 \wedge \ldots \wedge \bar{f}_m\ \Pi_1^k\, d\lambda$$

for $f_i \in F$, $i=1,\ldots,m$, $m=1,2,\ldots$ is equivalent to (c2). The equivalence of (c2) and (c2') follows from (4.2) since $\underline{f} \in K$ is the uniform limit of functions of the form $\bar{f}_1 \wedge \ldots \wedge \bar{f}_m$. ∎

Proof of Theorem 4.3. (Necessity) Let $f_i \in F$, $i=1,\ldots,m$. Then, since $P(\cdot|\underline{x}) \overset{U}{>} F_{\underline{x}}$, $\int f_i(y)P(dy|\underline{x}) \leq \bar{f}_i(\underline{x})$, and so,

$$\int f_1 \wedge\ldots\wedge f_m(y)P(dy|\underline{x}) \leq \bar{f}_1 \wedge\ldots\wedge \bar{f}_m(\underline{x}).$$

Thus,

$$\int f_1 \wedge\ldots\wedge f_m\, d\nu = \int\int f_1 \wedge\ldots\wedge f_m(y)P(dy|\underline{x}) \Pi_1^k\, d\lambda(x_i)$$

$$\leq \int \bar{f}_1 \wedge\ldots\wedge \bar{f}_m(\underline{x}) \Pi_1^k\, d\lambda(x_i).$$

(Sufficiency) Let $f \in C(K)$. Then, by Lemma 4.3, Corollary 4.5, Lemma 4.6 and the definition of ν,

$$\int f d\nu = \int \bar{f} d\nu_0 \leq \int h_f d\nu_0 \leq \int h_f(\underline{x}) \Pi_1^k\, d\lambda(x_i).$$

This with Theorem 1 of Strassen (1965) gives the representation of ν in terms of λ and a Markov kernel $P(\cdot|\underline{x}) \overset{U}{>} F_{\underline{x}}$. ∎

Remark 4.7. Let $\nu(A) = \int P(A|\underline{x}) \Pi_1^k\, d\lambda(x_i)$. Then $P(\cdot|\underline{x}) \overset{U}{>} F_{\underline{x}}$ for all $\underline{x}$ is equivalent to $E(\int f dF_{\underline{Y}}|\underline{X}) \geq \int f dF_{\underline{X}}$ where $X_1,X_2,\ldots,X_k$ are i.i.d. λ and, given $\underline{X} = \underline{x}$, $Y_1,\ldots,Y_k$ are i.i.d. $P(\cdot|\underline{x})$. When U is an ET-system an argument like that in Case 2 of Lemma 4.2 shows that this is equivalent to the martingale type of formula $E(\int u_j dF_{\underline{Y}}|\underline{X}) = \int u_j dF_{\underline{X}}$, $j=0,\ldots,2k-1$. For the classical ET-system $u_j(x) = x^j$, an apt name for a sequence $\underline{X}_1,\underline{X}_2,\ldots$ of random k-vectors satisfying

$$E(\int x^j dF_{\underline{X}_{n+1}}|\underline{X}_n,\ldots,\underline{X}_1) = \int x^j dF_{\underline{X}_n} \text{ for } j=0,\ldots, 2k-1$$ is a k-mart sequence. Theorem 4.1 characterizes the marginal p.m.'s that can correspond to a k-mart sequence.

References

Karlin, S. (1968). Total Positivity. Stanford University Press, Stanford, CA.

Karlin, S. and Studden, W. (1966). Tchebycheff Systems: With Applications in Analysis and Statistics. Interscience, NY.

Meyer, P. A. (1966). Probability and Potentials. Blaisdell, London.

Roberts, A.W. and Varberg, D.E. (1973). Convex Functions. Academic Press,NY.

Rockafellar, R. T. (1970). Convex Analysis. Princeton University Press, Princeton, NJ.

Schweder. T. (1982). On the dispersion of mixtures. Scand. J. Statist., 9, 165-169.

Shaked, M. (1980). On the mixtures from exponential families. J. R. Statist, B, 42, 192-198.

Strassen, V. (1965). The existence of probability measures with given marginals. Ann. Math. Statist., 36, 423-439.

END

DTIC

9-86