Total Positivity: A Review

by

Jee Soo Kim and Frank Proschan

Florida State University

FSU Statistics Report M663
AFSOR Technical Report No. 83-159

June, 1984

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

1Research Sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant (or cooperative agreement) Number AFOSR-82-K-0007. The U. S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

Key Words: Total Positivity, Pólya frequency function, multivariate total positivity of order 2, variation diminishing property, composition theorem, IFR, DFR, applications of total positivity.

American Math Society Subject Classification Nos.: 60-02, 62-02.
This paper is an invited entry for the Encyclopedia of Statistical Sciences, edited by N.L. Johnson and S. Kotz and published by John Wiley and Sons. The main objective is to review the concepts of total positivity, which play an important role in various domains of mathematics and statistics. This article describes the power and scope of total positivity, and samples the great variety of fields of its applications.
Total Positivity: A Review

by

Jee Soo Kim and Frank Proschan

ABSTRACT

This paper is an invited entry for the Encyclopedia of Statistical Sciences, edited by N. L. Johnson and S. Kotz and published by John Wiley and Sons. The main objective is to review the concepts of total positivity, which plays an important role in various domains of mathematics and statistics. This article describes the power and scope of total positivity, and samples the great variety of fields of its applications.

1. Introduction.

The theory of total positivity has been extensively applied in several domains of mathematics, statistics, economics, and mechanics. In statistics, totally positive functions are fundamental in permitting characterizations of best statistical procedures for decision problems. The scope and power of this concept extend to ascertaining optimal policy for inventory and system supply problems, to clarifying the structure of stochastic processes with continuous path functions, to evaluating the reliability of coherent systems, and to understanding notions of statistical dependency.

In recent years Samuel Karlin has made brilliant contributions in developing the intrinsic relevance and significance of the concept of total
positivity to probability and to statistical theory. In 1968, Karlin wrote a classical book devoted to this vast subject. This remarkable book presents a comprehensive, detailed treatment of the analytic structure of totally positive functions, and conveys the breadth of the great variety of fields of its applications. This book, together with Karlin's other fundamental papers, inspired many new developments and discoveries in many areas of statistical applications. Frydman and Singer encountered a complete solution to the embedding problem for the class of continuous-time Markov chains. The class of transition matrices for the finite state time-inhomogeneous birth and death processes coincides with the class of non-singular totally positive stochastic matrices. Keilson and Kester employed total positivity to characterize a class of stochastically monotone Markov chains which has the property that the expectation of unimodal functions of the chain is itself unimodal in the initial state. To help unify the area of stochastic comparisons Hollander, Proschan and Sethuraman introduced the concept of functions decreasing in transposition (DT). In the bivariate case, a function \( f(x_1, x_2; x_1, x_2) \) is said to have the DT property if

(a) \( f(x_1, x_2; x_1, x_2) = f(x_2, x_1; x_2, x_1) \) and
(b) \( x_1 < x_2, x_1 < x_2 \)

imply that \( f(x_1, x_2; x_1, x_2) \geq f(x_1, x_2; x_2, x_1) \); i.e., transposing from the natural order \( (x_1, x_2) \) to \( (x_2, x_1) \) decreases the value of the function. In their paper, total positivity is essential in showing that \( P_{x_1|x_2} \), the probability of rank order \( x_1 \), is a DT function.

Karlin and Rinott extended the theory to multivariate cases. Multivariate total positivity properties are instrumental in and for the results which are applied to obtain positive dependence of random vector components and related probability inequalities.
For an excellent global view of the theory, as well as for pertinent references, the reader may consult Karlin [15].

2. Definition and Basic Properties.

2.1. Definition of Totally Positive Function.

A function $f(x,y)$ of two real variables ranging over linearly ordered one-dimensional sets $X$ and $Y$, respectively, is said to be totally positive of order $k$ ($TP_k$) if for all $x_1 < x_2 < \ldots < x_m$, $y_1 < y_2 < \ldots < y_m$ ($x_i$ in $X$; $y_i$ in $Y$), and all $1 \leq m \leq k$,

$$f(x_1, y_1) \ f(x_1, y_2) \ \ldots \ f(x_1, y_m)$$
$$f(x_2, y_1) \ f(x_2, y_2) \ \ldots \ f(x_2, y_m)$$
$$\vdots \quad \vdots \quad \ddots \quad \vdots$$
$$f(x_m, y_1) \ f(x_m, y_2) \ \ldots \ f(x_m, y_m)$$

$$\geq 0.$$  \hspace{1cm} (1)

Typically, $X$ and $Y$ are either intervals of the real line or a countable set of discrete values on the real line, such as the set of all integers or the set of nonnegative integers. When $X$ or $Y$ is a set of integers, the term "sequence" rather than "function" is used. If $f(x,y)$ is $TP_k$ for all positive integers $k=1, 2, \ldots$, then $f(x,y)$ is said to be totally positive of order $\infty$, written $TP_{\infty}$ or $TP$.

A related, weaker property is that of sign regularity. A function $f(x,y)$ is sign regular of order $k$ ($SR_k$) if for every $x_1 < x_2 < \ldots < x_m$, $y_1 < y_2 < \ldots < y_m$, and $1 \leq m \leq k$, the sign of

$$\begin{vmatrix}
  x_1, x_2, \ldots, x_m \\
  y_1, y_2, \ldots, y_m 
\end{vmatrix}$$

depends on $m$ alone.
Many well known families of density functions (both continuous and discrete) are totally positive. It should be noted that $TP_2$ is the order of TP-ness which has found greatest application. In the context of statistics, the $TP_2$ property is referred to as the monotone likelihood ratio property. Higher order TP-ness is hardly used in application except for the occasional use of $TP_3$.

Some examples of functions that possess the TP property are:

(i) $f(x,y) = e^{xy}$ is TP in $x, y \in (-\infty, \infty)$, so that $f(x,y) = x^y$ is TP in $x \in (0, \infty)$ and $y \in (-\infty, \infty)$.

(ii) $f(k,t) = e^{-t} [ (kt)^k / k! ]$ is TP in $t \in (0, \infty)$ and $k \{0, 1, 2, \ldots\}$.

(iii) $f(x,y) = \begin{cases} 1 \text{ if } a \leq x \leq y \leq b \\ 0 \text{ if } a \leq y \leq x \leq b \end{cases}$

2.2. $PF_k$ as Special Case of Interest.

The concepts of $TP_1$ and $TP_2$ densities are familiar ones. Every density is $TP_1$; while the $TP_2$ densities are those having a monotone likelihood ratio.

A further important specialization occurs if a $TP_k$ function may be written as a function $f(x,y) = f(x-y)$ of the difference of $x$ and $y$, where $x$ and $y$ traverse the entire real line; $f(u)$ is then said to be a Pólya frequency function of order $k(PF_k)$. Note that a Pólya frequency function is not necessarily a probability frequency function in that $\int_{-\infty}^{\infty} f(u)du$ need not be 1 nor even finite.

The class of $PF_2$ functions is particularly important and has rich applications to decision theory [10], [11], [12], [18], reliability theory [5], and the stochastic theory of inventory control models [1], [16].

Every $PF_2$ function is of the form $e^{-\psi(x)}$, where $\psi(x)$ is convex. On the other hand, there exists no such simple representation for $PF_k$, $k \geq 3$. Probability densities which are $PF_2$ abound. For other properties and examples
Probability densities which decrease to zero at an algebraic rate in the tails are not PF. For example, (i) Weibull with shape parameter $\alpha < 1$: $f(x) = \alpha \lambda x^{\alpha-1} \exp[-(\lambda x)^\alpha]$, $x \geq 0$, $\lambda > 0$, $0 < \alpha < 1$, and (ii) Cauchy: $f(x) = 1/[\pi(1+x^2)]$, $-\infty < x < \infty$ are not PF.

Intriguing results in the structure theory of PF functions can be found in Karlin and Proschan [16], Karlin, Proschan, and Barlow [17], and Barlow and Marshall [2].

2.3. Variation Diminishing Property.

An important feature of totally positive functions of finite or infinite order is their variation diminishing property: If $f(x,y)$ is TP and $g(y)$ changes sign at most $j \leq k-1$ times, then $h(x) \int f(x,y) g(y) \, dy$ changes sign at most $j$ times; moreover, if $h(x)$ actually changes sign $j$ times, then it must change sign in the same order as $g(y)$. It is this distinctive property which makes TP so useful. The variation diminishing property is essentially equivalent to the determinantal inequalities (1). Greater generality in stating this property is possible. The interested reader is referred to Chapter 1, Karlin [13]. A more direct approach to the theory is taken by Brown et al. [5], giving appropriate definitions and criteria for checking directly whether a family of densities possesses variation diminishing property.

2.4. Composition and Preservation Properties.

Many of the structural properties of TP functions are deducible from the following basic identity which is an indispensible tool in the study of total positivity.

**Basic Composition Formula.** Let $h(x,t) = \int f(x,y) g(y,t) \, dy \, \sigma(y)$ converge abso-
lutely, where \( d \sigma (y) \) is a sigma-finite measure. Then

\[
h \left( \begin{array}{c} x_1, x_2, \ldots, x_n \\ t_1, t_2, \ldots, t_n \end{array} \right) = \int \cdots \int f \left( \begin{array}{c} x_1, x_2, \ldots, x_n \\ y_1, y_2, \ldots, y_n \end{array} \right) \left[ \begin{array}{c} y_1, y_2, \ldots, y_n \\ t_1, t_2, \ldots, t_n \end{array} \right] d \sigma (y_1) \ldots d \sigma (y_n). \tag{2} \]

A direct consequence of the composition formula is: If \( f(x,y) \) is TP and \( g(y,t) \) is TP, then

\[
h(x,t) = \int f(x,y) g(y,t) d \sigma (y) \]

is TP \( \min(m,n) \). In many statistical applications this consequence is exploited principally in the case when \( f \) and \( g \) are Pólya frequency densities. That is, if \( f(x) \) is PF and \( g(x) \) is PF, then

\[
h(x) = \int f(x-t) g(t) dt \]

is PF \( \min(m,n) \). From this we can obtain a key result as follows.

**Theorem 1.** Let \( f_1, f_2, \ldots \) be density functions of nonnegative random variables with each \( f_i \) a PF. Then

\[
g(n,x) = f_1 \ast f_2 \ast \cdots \ast f_n (x) \]

(* indicates convolution) is PT in the variables \( n \) and \( x \), where \( n \) ranges over 1, 2, \ldots and \( x \) traverses the positive real line.

The case when the random variables are not restricted to be nonnegative is discussed in Karlin and Proschan [16]. These composition and preservation properties allow us to generate other totally positive functions, thus making it easy to enlarge the TP or PF classes and to determine whether the TP property holds.

### 2.5. Unimodality and Smoothness Properties

A function totally positive or more generally sign regular is endowed with certain structural properties pertaining to unimodality and smoothing properties. From the definition of PF \( 2 \) can be derived
\[
\begin{vmatrix}
  f(x_1-y) & -f'(x_1-y) \\
  f(x_2-y) & -f'(x_2-y)
\end{vmatrix} \geq 0
\] (3)

for \( x_1 < x_2 \) and \( y \) arbitrary.

In the event that \( f'(u_0) = 0 \) the above inequality implies that \( f'(u) \geq 0 \) for \( u < u_0 \) and \( f'(u) \leq 0 \) for \( u > u_0 \). This clearly implies that if \( f(u) \) is PF \( \text{ and } f(u) \) is unimodal. In particular, every \( \text{PF} \) density is a unimodal density.

We note that the unimodality result is valid in case \( f \) is a \( \text{PF} \) sequence.

We now describe a smoothening property possessed by the transformation under which convexity in \( g(x) \) is carried over into convexity in \( h(x) \), viz.

\[
h(n) = \int f^{(n)}(x)g(x)dx \quad \text{for } n = 1, 2, \ldots
\] (4)

where \( f^{(n)}(x) \) is the \( n \)-fold convolution of \( f \). To make this notion precise assume \( f(x) \) is \( \text{PF} \) and \( g(x) \) is convex. Let \( u = \int xf(x)dx \). Note that for arbitrary real constants \( a_0 \) and \( a_1 \),

\[
\int \{ g(x) - [(a_0/u)x+a_1] \} f^{(n)}(x)dx = h(n) - (a_0n+a_1).
\] (5)

Since \( g(x) \) is convex, then \( g(x) - [(a_0/u)x+a_1] \) has at most 2 changes of sign and if 2 changes of sign actually occur, they occur in the order + - + as \( x \) traverses the real axis from -\( \infty \) to +\( \infty \). Since \( f \) is \( \text{PF} \), \( f^{(n)}(x) \) is \( \text{TP} \) in the variables \( n \) and \( x \) by Theorem 1. The variation diminishing property implies that \( h(n) - (a_0n+a_1) \) will have at most 2 changes of sign. Moreover, if \( h(n) - (a_0n+a_1) \) has exactly 2 changes of sign, then these will occur in the same order as those of \( g(x) - [(a_0/u)x+a_1] \), namely + - +. Since \( a_0, a_1 \) are arbitrary, we easily infer that \( h(n) \) is a convex function of \( n \). Similar results apply for concavity.
Applications to Statistical Decision Theory.

Historically this is perhaps the first area of statistics benefiting from the application of TP due to the great papers of Karlin [10, 11, 12]. We consider the problem of testing a null hypothesis against its alternative hypothesis, i.e., a 2-action statistical decision problem. There exist two loss functions $L_1$ and $L_2$ on the parameter space where $L_i(\theta)$ is the loss incurred if action $i$ is taken when $\theta$ is the true parameter value. The set in which $L_1(\theta) < L_2(\theta)$ is the set in which action 1 (action 2) is preferred when $\theta$ is the true state of nature. The two actions are indifferent at all other points. We shall assume that $L_1(\theta) - L_2(\theta)$ changes sign exactly $n$ times at $\theta_1, \theta_2, \ldots, \theta_n$.

Let $\phi$ be a randomized decision procedure which is the probability of taking action 2 (accepting the alternative hypothesis) if $x$ is the observed value of the random variable $X$. Let $C_n$ be the class of all monotone randomized decision procedures defined by

$$
\phi(x) = \begin{cases} 
1 & \text{for } x_{2i} < x < x_{2i+1}, \ i = 0, 1, \ldots, \left[\frac{n}{2}\right] \\
\frac{i}{j} & \text{for } x = x_j, \ 0 \leq i_j \leq 1, \ j = 1, 2, \ldots, n \\
0 & \text{elsewhere,}
\end{cases}
$$

(6)

where $[a]$ denotes the greatest integer $\leq a$ and $x_0 = -\infty$.

Using the variation diminishing property, Karlin [11] obtained the main results, which state:

**Theorem 2.** Let $f(x, \theta)$ be a strictly $TP_{n+1}$ density and $\phi(\theta, \phi) = \int [(1 - \phi(x)) L_1(\theta) + \phi(x) L_2(\theta)] f(x, \theta) \, dx(x)$. Then for any randomized decision procedure $\phi$ not in $C_n$ there exists a unique $\phi^0$ such that $f(\theta, \phi^0) \leq f(\theta, \phi)$ with inequalities everywhere except for $\theta = \theta_1, \theta_2, \ldots, \theta_n$. 

Theorem 3. If $\varphi$ and $\psi$ are two procedures in $C_n$ and $f$ is strictly TP$_{n+1}$ then
\[ \int \varphi(x) \psi(x) f(x,\omega) \, d\omega(x) \] has less than $n$ zeros counting multiplicities.
Assume $f(x,\omega)$ is strictly TP$_2$. For a one-sided testing problem, existence of a uniformly most powerful level $\alpha$ test can be easily established by Theorem 2 and Theorem 3.
More detailed discussions and other decision theoretic applications can be found in Karlin [10], [11], [12] and Karlin and Rubin [20].


Let $P(t,x,E)$ be the transition probability function of a homogeneous strong Markov process whose state space is an interval on the real line and which possesses a realization in which almost all sample paths are continuous.
Karlin and McGregor [14] established the intimate relationship between the general theory of TP functions and the theory of diffusion stochastic processes. Their main result shows the transition probability function $P(t,x,E)$ is totally positive in variables $x$ and $E$. That is, if $x_1 < x_2 < \ldots < x_n$ and $E_1 < E_2 < \ldots < E_n$ (where $x < y$ for every $x \in E_i$ and $y \in E_j$), then $\det|P(t,x_i,E_j)| > 0$ for every $t > 0$ and integer $n$. This relation introduces the concept of a TP set function $f(x,E) = P(t,x,E)$ where $t$ is fixed, $x$ ranges over a subset of the real line, and $E$ is a member of a given sigma field of sets on the line.

If the state space of the process is countably discrete, then continuity of the path functions means that in every transition of the process the particle changes "position", moving to one of its neighboring states. Thus, discrete state continuous path processes coincide with the so-called birth-death processes (Karlin and McGregor [15]) which is a stationary Markov process whose transition probability matrix $P_{ij}(t) = Pr(x(t) = j \mid x(0) = i)$ is totally positive in the values $i$ and $j$ for every $t > 0$. 
Two concrete illustrations of transition probability functions that arise from suitable diffusion processes are [14]:

(i) Let $L^2_n(x)$ be the usual Laguerre polynomial, normalized so that $L^2_n(0) = (n+1)^{n+1} / n!$, and let $P(t)$ be the infinite matrix with elements

$$ P_{mn}(t) = \int_0^\infty e^{-x} L^2_n(x) L^2_m(x) x^{n-1} e^{-x} \, dx. $$

Then $P(t)$ is strictly TP for each fixed $t > 0$ and $x > -1$. This is an example of a transition probability matrix for a birth-death process.

(ii) The Wiener process on the real line is a strong Markov process with continuous path functions. The direct product of $n$ copies of this process is the $n$-dimensional Wiener process which is known to be a strong Markov process. Therefore the transition probability function $P(t,x,y) = 1/(1+4t) \exp[-(x-y)^2/4t] dy$ is totally positive for $t > 0$.

5. Applications in Inventory Problem.

Suppose that the probability density $f(x)$ of demand for each period is a PF. The policy followed is to maintain the stock size at a fixed level $S$ which will be suitably chosen so as to minimize appropriate expected costs, or is determined by a fixed capacity restriction. At the end of each period an order is placed to replenish the stock consumed during that period so that a constant stock level is maintained on the books. Delivery takes place $n$ periods later. The expected cost for a stationary period as a function of the lag is

$$ L(n) = \int_0^S h(S-y)f^{(n)}(y) dy + \int_0^\infty \phi(y-S)f^{(n)}(y) dy $$

where $S$ is fixed, $h$ represents the storage cost function and $\phi$ the penalty cost function.

Let $h$ and $\phi$ be convex increasing functions with $h(0) = \phi(0) = 0$. Then
we may write

\[ L(n) = \int_0^S h(S-y) \, dy + \int_S^\infty f(n)(y) \, dy, \]

where

\[ f(y) = \begin{cases} g(y-S) & \text{for } S < y, \\ f_n(y) & \text{for } 0 \leq y \leq S. \end{cases} \]

Now \( f(y) \) is a convex function. Hence by the convexity preserving property of (4), we conclude that \( L(n) \) is a convex function. Thus, if the length of \( h \) increases, the marginal expected loss increases.

Very interesting applications of total positivity are found in system supply problems. Suppose we wish to determine the intitial spare-parts kit for a complex system which provides maximum assurance against system shutdown due to shortage of essential components during a period of length \( t \) under a budget for spares \( C_0 \). We assume that a failed component is instantly replaced by a spare, if available. Only spares originally provided may be used for replacement, i.e., no resupply of spares can occur during the period. The system contains \( d_i \) operating components of type \( i \), \( i = 1, 2, \ldots, k \). The length of life of the \( j^{th} \) operating component of the \( i^{th} \) type is assumed to be an independent random variable with PDF \( g_{ij} \), \( j = 1, 2, \ldots, d_i \). The unit cost of a component of type \( i \) is \( c_i \).

Our problem is to find \( n_i \), the number of spares initially provided of the \( i^{th} \) type, such that \( \sum_{i=1}^k P_i(n_i) \) is maximized subject to \( \sum_{i=1}^k n_i c_i \leq C_0 \) and \( n_i = 0, 1, 2, \ldots \) for \( i = 1, 2, \ldots, k \), where \( P_i(m) = \text{probability of experiencing } \leq m \text{ failures of type } i \).

In Black and Proschan [4], a detailed discussion of methods is given for computing the solution when each \( \ln P_i(m) \) is concave in \( m \), or equivalently, when each \( P_i(n-m) \) is a TP-sequence in \( n \) and \( m \). To show \( P_i(n-m) \) is a TP-sequence in \( n \) and \( m \), we note:
1. \( c_{ij}(n) \), the probability of requiring \( n \) replacements of operating component \( i, j \), is a PF\(_2\) sequence in \( n \) for each fixed \( i \) and \( j \).

2. \( c_i(n) \), the probability of requiring \( n \) replacements of type \( i \), is a PF\(_2\) sequence in \( n \) for each \( i \), since \( c_i(n) = c_{i1} \ast c_{i2} \ast \ldots \ast c_{id_i}(n) \).

3. \( P_i(n-m) \) is a TP\(_2\) sequence in \( n \) and \( m \) for each \( i \), since

\[
\begin{align*}
(a) \quad P_i(n) &= \sum_{m=0}^{\infty} c_i(n-m)q(m), \\
q(m) &= \begin{cases} 
1 & \text{for } m = 0, 1, 2, \ldots \\
0 & \text{otherwise},
\end{cases}
\end{align*}
\]

(b) \( q(m) \) is a PF\(_m\) sequence,

and

(c) the convolution of PF\(_k\) is PF\(_k\).

A procedure for computing the optimal spare parts kit in terms of \( P_i(n-m) \) is given in [4]: For arbitrary \( r > 0 \), for those \( i \) such that \( \sum n_i(n+1) - \sum n_i(n) \geq rc_i \), define \( n^*_i(r) = 0 \); for the remaining \( i \), define \( n^*_i(r) \) as the largest \( n \) such that \( \sum n_i(n+1) - \sum n_i(n) \geq rc_i \). Compute \( c_n^*(r) = \sum_{i=1}^{k} c_i n^*_i(r) \).

\( n^* \) is optimal when \( c_0 \) is once of the values assumed by \( c_n^*(r) \) as \( r \) varies over \((0, \infty)\).

6. Applications in Reliability and Life Testing.

A life distribution \( F \) is said to have increasing (decreasing) failure rate, denoted by IFR (DFR), if \( \log [1-F(t)]/\log \hat{F}(t) \) is concave (convex on \((0, \infty)\)). If \( F \) has a density \( f \), then the failure rate at time \( t \) is defined by \( r(t) = f(t)/\hat{F}(t) \) for \( F(t) < 1 \). Distributions with monotone failure rate are of considerable practical interest and such distributions constitute a very large class.
The monotonicity properties of the failure rate function \( r(t) \) are intimately connected with the theory of total positivity. The statement that a distribution \( F \) has an increasing failure rate is equivalent to the statement that \( F(x) - F(y) \) is TP in \( x \) and \( y \), or \( F(x) \) is PF.

The concept of TP yields fruitful applications in shock models. We say that a distribution \( F \) has increasing failure rate average (IFRA) if \( (1/t) \cdot \log F(t) \) is increasing in \( t > 0 \), or equivalently, \( F(t) \cdot t \) is decreasing in \( t > 0 \). An IFRA distribution provides a natural description of coherent system life when system components are independent HIR. The IFRA distribution also arises naturally when shocks occur randomly according to a Poisson process with intensity \( \lambda \). The \( i \)th shock causes a random amount \( X_i \) of damage, where \( X_1, X_2, \ldots \) are independently distributed with common distribution \( F \).

A device fails when the total accumulated damage exceeds a specified capacity or threshold \( x \). Let \( \bar{H}(t) \) denote the probability that the device survives \( 0, t \). Then

\[
\bar{H}(t) = \begin{cases} 
\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda t} k! / k! & \text{for } 0 < t < \infty \\
1 & \text{for } t \leq 0.
\end{cases}
\]

Note that \( e^{-\lambda t} (\lambda t)^k / k! \) represents the Poisson probability that the device experiences exactly \( k \) shocks in \((0,t]\), while \( \bar{F}_k = F^{(k)}(x) \) represents the probability that the total damage accumulated over the \( k \) shocks does not exceed the threshold \( x \), with \( 1 = \bar{F}_0 \geq \bar{F}_1 \geq \bar{F}_2 \geq \ldots \).

As key tools in deriving the main result in shock models, the methods of total positivity are employed and in particular the variation diminishing property of TP functions.
If \( \bar{F}_k^{1/k} \) is decreasing in \( k \), \( \bar{F}_k \), \( 0 \leq \xi \leq 1 \), has at most one sign change from + to - if one occurs. Then it follows from the variation diminishing property that \( (\bar{H}(t) \bar{t})^{1/t} \) is decreasing in \( t \), i.e., \( H \) is IFRA.

The following implications are readily checked:

- \( \text{PF}_1 \) density \( \rightarrow \) IFR distribution \( \rightarrow \) IFRA distribution.

For further discussion and illustrations of the usefulness of total positivity in reliability practices we refer to Barlow and Proschan [3].

### 7. Multivariate Total Positivity and its Relationship to Qualitative Notions of Dependency

The following natural generalization of \( \text{TP}_2 \) was introduced and studied by Karlin and Rinott [18].

**Definition.** Consider a function \( f(x) \) defined on \( \mathbb{X} = X_1 \times X_2 \times \ldots \times X_k \) where each \( X_i \) is totally ordered. We say that \( f(x) \) is multivariate totally positive of order \( 2 \) or \( \text{MTP}_2 \) if

\[
f(x \vee y) f(x \wedge y) \leq f(x) f(y) \quad \text{for every } x, y \in \mathbb{X},
\]

where \( x \vee y = (\max(x_1, y_1), \max(x_2, y_2), \ldots, \max(x_k, y_k)) \) and

\( x \wedge y = (\min(x_1, y_1), \min(x_2, y_2), \ldots, \min(x_k, y_k)) \). In order to verify (11) it suffices to show that \( f(x) > 0 \) is \( \text{TP}_2 \) in every pair of variables where the remaining variables are held fixed.

Multivariate normal distributions constitute an important class of \( \text{MTP}_2 \) probability densities. Let \( \mathbf{X} \) follow the density

\[
f(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-u)\Sigma^{-1}(x-u)^T\right),
\]

where \( \Sigma^{-1} = B = \sum b_{ij}^2 i, j=1 \). This density is \( \text{TP}_2 \) in each pair of arguments and hence \( \text{MTP}_2 \) if and only if \( b_{ij} \leq 0 \) for all \( i \neq j \).
In a great many situations, the random variables of interest are not independent. To appropriately model these situations, Esary, Proschan and Walkup introduced the concept of association of random variables. Random variables \( X_1, X_2, \ldots, X_n \) are said to be associated if \( \text{Cov}(f(X), g(X)) \geq 0 \) for all pairs of increasing functions \( f \) and \( g \).

It can be shown that if \( \bar{X} = (X_1, X_2, \ldots, X_n) \) has a joint MTP\(_2\) density, the \( E:1(\bar{X}) \leq E:1(\bar{X}') \) holds provided \( \bar{X} \) and \( \bar{X}' \) are simultaneously monotone increasing (or decreasing). Equivalently, \( \text{Cov}(X_i(X), \bar{X}j(X)) \geq 0 \). Thus an MTP\(_2\) random vector \( \bar{X} \) consists of associated random variables.

It is well known that the union of independent sets of associated random variables produces an enlarged set of associated random variables. Clearly increasing functions of associated random variables are again associated. It follows that if \( \bar{X} \) and \( \bar{Y} \) are independent random variables each with associated components, then the components of \( \bar{Z} = \bar{X} + \bar{Y} \) is associated. Thus, in particular, if \( \bar{X} \) and \( \bar{Y} \) both have MTP\(_2\) densities, then association of \( (Z_1, Z_2, \ldots, Z_n) \) is retained. However, \( \bar{Z} \) need not have a joint MTP\(_2\) density.

A key to many of the results on positive dependence and probabilistic inequalities for the multinormal, multivariate t, and Wishart distributions obtained by Karlin and Rinott is the degree of MTP\(_2\) property inherent in these distributions. Their main theorem delineates a necessary and sufficient condition that the density of \( \bar{X} = (|X_1|, |X_2|, \ldots, |X_n|) \) where \( X = (X_1, X_2, \ldots, X_n) \) is governed by \( X \sim N(0, \Sigma) \) be MTP\(_2\) is that there exists a diagonal matrix \( D \) with diagonal elements \( l_i \) such that the off-diagonal elements of \( -D \Sigma^{-1} D \) are all nonnegative. For an illustration of the power of this theorem consider \( \bar{X} = (|X_1|, |X_2|, \ldots, |X_n|) \) possessing a joint MTP\(_2\) density where

\[
X \sim N(0, \Sigma), \quad \text{Define} \quad S_i = \sum_{j=1}^{n} X_{ij}^2, \quad i = 1, 2, \ldots, n, \quad \text{where}
\]

\[
X_{\nu} = (X_{\nu 1}, X_{\nu 2}, \ldots, X_{\nu n}), \quad \nu = 1, 2, \ldots, p \text{ are i.i.d. random vectors}
\]
satisfying the condition of the theorem. The random variables $S_1, S_2, \ldots, S_n$ are associated and have the distribution of the diagonal elements of a random positive definite $n \times n$ matrix $S$ where $S$ follows the Wishart distribution $W_n(p, \Sigma)$ with $p$ degrees of freedom and parameter $\Sigma$. It is established in [19] that $P_i(S_i \geq c_i, S_2 \geq c_2, \ldots, S_n \geq c_n) \geq \prod_{i=1}^{n} P_i(S_i \geq c_i)$ for any positive $c_i$. For other applications and ramifications of MTP, see Karlin and Rinott [18], [19].

Fahmy et al. [7] exploited the concept of MTP to obtain interesting results on assessing the effect of the sample on the posterior distribution in the Bayesian context.


