ON THE STRUCTURE OF $2 \times \infty$ BIVARIATE DISTRIBUTIONS WHICH ARE TOTALLY POSITIVE OF ORDER TWO

BY K. SUBRAMANYAM[1] AND M. BHASKARA RAO[1]

University of North Carolina at Wilmington and North Dakota State University

Let X and Y be two random variables such that X takes only two values 1 and 2. The notion of total positivity of order two for the joint probability distribution of X and Y is discussed in this paper from the viewpoint of convex analysis. The set of all $2 \times \infty$ probability measures which are totally positive of order two and with fixed second marginal probability measure is shown to be convex. Some of the extreme points of this set are explicitly spelled out, and an integral representation theorem in terms of extreme points is presented in a special case.

1. Introduction. Let X and Y be two random variables having a joint probability density function $f(\cdot,\cdot)$ with respect to some product probability measure λ on the Borel σ-field of R^2. The random variables X and Y are said to be totally positive of order two if the determinants

$$\begin{vmatrix} f(x,y) & f(x,y') \\ f(x',y) & f(x',y') \end{vmatrix}$$

are nonnegative for $-\infty < x \leq x' < \infty$ and $-\infty < y \leq y' < \infty$ a.e. $[\lambda]$. See Karlin (1968, p. 12). For its relation with other notions of dependence and further ramifications, see Barlow and Proschan (1981). See also Lehmann (1966).

The main purpose of this article is to perform extreme point analysis on the notion of total positivity of order two. What this means is that we look at the set of all bivariate probability density functions, examine convexity of this set, and if convex, enumerate all its extreme points. This kind of analysis was carried out on a limited scale in Subramanyam and Bhaskara Rao (1988). It was shown that the set of all bivariate probability density functions which are totally positive of order

[1]Supported by AFOSR contract AFSO–88–0030.

AMS 1980 subject classifications. Primary 60E05; secondary 62H05.

Key words and phrases. Bivariate distribution, total positivity of order two, convex set, compact set, extreme points.

The authors are very much grateful to the referee for many of his perceptive comments which led to a substantial improvement in the presentation of the paper.

two is not convex. The attention then was focused on the set of all $2 \times n$ bivariate distributions

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \end{pmatrix}$$

which are totally positive of order two and with fixed column marginals $p_{11}+p_{21} = q_1$, $p_{12} + p_{22} = q_2, \ldots, p_{1n} + p_{2n} = q_n$. The extreme points of this convex set were explicitly enumerated. The analysis of this convex set was found to be useful in testing certain hypotheses of independence and total positivity of order two.

The main thrust of this paper is in analyzing total positivity of order two in the realm of $2 \times \infty$ bivariate distributions. In Section 2, the set $M_\mu(\mathrm{TP}_2)$ of all $2 \times \infty$ probability measures which are totally positive of order two and with fixed second marginal probability measure μ is shown to be convex. Some of the extreme points of the set $M_\mu(\mathrm{TP}_2)$ are explicitly spelled out, and an integral representation of any given λ in $M_\mu(\mathrm{TP}_2)$ in terms of extreme points is presented in a special case. Some open questions are raised on the extreme points of the set $M_\mu(\mathrm{TP}_2)$.

2. Main Results. To begin with, we frame the definition of TP_2 in the language of probability measures. The basic notation is as follows. Let $\Omega = \{1,2\} \times R$. The space Ω consists of two lines $x = 1$ and $x = 2$ in R^2.

Let $\mathcal{C}$ be the Borel σ-field on R. We equip Ω with the following σ-field.

$$\mathcal{B} = \{A \subset \Omega;\ A = \{1\} \times B_1 \cup \{2\} \times B_2 \text{ for some } B_1 \text{ and } B_2 \text{ in } \mathcal{C}\}.$$

The above representation of A is unique.

For any probability measure μ on $\mathcal{B}$, let μ_1 and μ_2 denote the first and second marginal probability measures of μ on $\{1,2\}$ and $\mathcal{C}$, respectively, i.e.,

$$\begin{aligned} \mu_1(\{1\}) &= \mu(\{1\} \times R), \\ \mu_1(\{2\}) &= \mu(\{2\} \times R), \end{aligned}$$

and

$$\mu_2(B) = \mu(\{1,2\} \times B), \quad B \in \mathcal{C}.$$

For any two probability measures τ and ν on $\{1,2\}$ and $\mathcal{C}$, respectively, let $\tau \otimes \nu$ denote the product probability measure on $\mathcal{B}$. The probability measure $\tau \otimes \nu$ has the following explicit formula. For any $A = \{1\} \times B_1 \cup \{2\} \times B_2$ in $\mathcal{B}$ with B_1, $B_2 \in \mathcal{C}$,

$$(\tau \otimes \nu)(A) = \tau(\{1\})\nu(B_1) + \tau(\{2\})\nu(B_2).$$

For any two probability measures μ and λ, we use the notation $\mu << \lambda$ if μ is absolutely continuous with respect to λ.

For basic ideas on absolute continuity and product measures, see Halmos (1950).

DEFINITION. A probability measure μ on $\mathcal{B}$ is said to be totally positive of order two (TP_2) if the following determinants

$$D(y, y') = \begin{vmatrix} f(1, y) & f(1, y') \\ f(2, y) & f(2, y') \end{vmatrix}$$

are nonnegative almost surely $-\infty < y < y' < \infty$, where f is a version of the Radon-Nikodym derivative of μ with respect to some product probability measure $\tau \otimes \nu$ on $\mathcal{B}$ for which $\mu << \tau \otimes \nu$.

Some comments are in order on the above definition.

1. In the parlance of statistical theory, f is called a probability density function. One may wonder why one needs the dominating measure $\tau \otimes \nu$ to be a product measure in the above definition. If we were to allow any measure to dominate μ so as to get a density function, we could as well take μ itself as the dominating measure which gives the density function $f \equiv 1$. Then μ is TP_2 always! For the above definition to be nontrivial, we need to take the dominating measure to be a product measure. Moreover, the idea that μ is TP_2 is a deviation from independence, and to facilitate to measure the extent of deviation from independence one has to incorporate a product probability measure in the definition of TP_2. Thus a product probability measure enters the definition of TP_2 in the form of a dominating measure.

2. The statement that the determinants $D(y, y')$ are nonnegative almost surely $-\infty < y < y' < \infty$ requires some explanation. Let

$$U = \{(y, y') : -\infty < y < y' < \infty\}.$$

Let ν_U be the probability measure on the Borel σ-field of U defined by

$$\nu_U(A) = [\nu \otimes \nu(A)]/[\nu \otimes \nu(U)]$$

for every Borel subset A of U. If we let $A = \{(y, y') \,\varepsilon\, U : D(y, y') \geq 0\}$, then the TP_2 condition is equivalent to $\nu_U(A) = 1$.

3. It is assumed that $0 < \mu(\{1\}) < 1$ since, otherwise, μ is trivially TP_2.

4. The definition that μ being TP_2 can be rephrased purely in terms of μ dispensing totally with the necessity of working with probability density functions. The following is a result in that direction.

THEOREM 1. *A probability measure μ on $\mathcal{B}$ is TP_2 if and only if*

$$\mu(\{1\} \times [a, b])\mu(\{2\} \times [c, d]) \geq \mu(\{1\} \times [c, d])\mu(\{2\} \times [a, b]) \quad \text{for all } -\infty < a \leq b < c \leq d < \infty. \tag{1}$$

Block, Savits, and Shaked (1982) frame the definition of TP_2 for probability measures. Their remarks (iii) and (iv) on page 767 are more or less tantamount to the statement of the above theorem. We will not give a proof of this result here.

REMARK. In the above theorem, one can have either open intervals or semi-open intervals in (1). In the terminology of random variables, the notion of TP_2 has the following description. Let X and Y be two random variables such that X takes values 1 and 2. Then X and Y are TP_2 if

$$\begin{aligned} &P(X=1,\ a<Y\leq b)P(X=2,\ c<Y\leq d)\\ &\quad\geq P(X=1,\ c<Y\leq d)P(X=2,\ a<Y\leq b)\\ &\quad\text{for all } -\infty<a\leq b<c\leq d<\infty. \end{aligned}$$

This implies that

$$P(X=1,\ Y\leq y)/P(X=2,\ Y\leq y)$$

is a decreasing function of y.

5. A natural product probability measure dominating μ is $\tau\otimes\mu_2$, where τ is a nontrivial measure on $\{1,2\}$. Since all such product measures are mutually absolutely continuous, it will be convenient for us to let $\tau(\{1\})=\frac{1}{2}=\tau(\{2\})$.

Convexity Property. The set of all probability measures on $\mathcal{B}$ each of which is TP_2 is not convex. Examples are easy to construct. See Subramanyam and Bhaskara Rao (1988). We look at the following subset. Let ν be a fixed probability measure on $\mathcal{C}$. Let $M_\nu(TP_2)$ be the set of all probability measures μ on $\mathcal{C}$ such that μ is TP_2 and $\mu_2=\nu$. We confine our attention to TP_2 measures whose second marginal is a fixed probability measure ν. For all μ in $M_\nu(TP_2)$, we take Radon-Nikodym derivatives with respect to the fixed product probability measure $\tau\otimes\nu$, where $\tau(\{1\})=\frac{1}{2}=\tau(\{2\})$.

We now study some of the properties of $M_\nu(TP_2)$.

PROPOSITION 1. *Let $\mu\in M_\nu(TP_2)$. Let f be a version of the Radon-Nikodym derivative of μ with respect to $\tau\otimes\nu$. Then*

$$\frac{1}{2}f(1,y)+\frac{1}{2}f(2,y)=1 \text{ for almost all } y\ [\nu].$$

PROOF. Observe that for every B in $\mathcal{C}$,

$$\begin{aligned}
\nu(B) &= \mu_2(B) = \mu(\{1,2\} \times B) \\
&= \int_{\{1,2\}\times B} f(x,y)(\tau \otimes \nu)(d(x,y)) \\
&= \int_{\{1,2\}} \int_B f(x,y)\tau(dx)\nu(dy) \\
&= \int_B \left(\frac{1}{2} f(1,y) + \frac{1}{2} f(2,y)\right) \nu(dy).
\end{aligned}$$

From this, the proposition follows:

PROPOSITION 2. *Let $\mu \in M_\nu(TP_2)$ and f a version of the Radon-Nikodym derivative of μ with respect to $\tau \otimes \nu$. Then*

(i) $f(1,y)$ is a decreasing function of y almost surely $[\nu_U]$

and

(ii) $f(2,y)$ is an increasing function of y almost surely $[\nu_U]$.

PROOF. Since $\mu \in M_\nu(\mathrm{TP}_2)$, $f(1,y)f(2,y') \geq f(1,y')f(2,y)$ a.s. $[\nu_U]$. Thus,

$$f(1,y)[1 - \frac{1}{2} f(1,y')] \geq f(1,y')[1 - \frac{1}{2} f(1,y)] \text{ a.s. } [\nu_U]$$

and so, $f(1,y) \geq f(1,y')$ a.s. $[\nu_U]$. Thus, (i) follows. (ii) is a consequence of (i) and Proposition 1.

THEOREM 2. *The set $M_\nu(TP_2)$ is a compact convex set. (Compactness is in the topology of weak* convergence.)*

PROOF. We first settle convexity. Let μ and λ belong to $M_\nu(\mathrm{TP}_2)$ and $0 \leq \alpha \leq 1$. Let f and g be versions of Radon-Nikodym derivatives of μ and λ, respectively, with respect to $\tau \otimes \nu$. Observe that $\alpha f + (1-\alpha)g$ is a version of the Radon-Nikodym derivative of $\alpha\mu + (1-\alpha)\lambda$ with respect to $\tau \otimes \nu$. Then a.s. $[\nu_U]$,

$$\begin{aligned}
& [\alpha f(1,y) + (1-\alpha)g(1,y)][\alpha f(2,y') + (1-\alpha)g(2,y'] \\
- & [\alpha f(1,y') + (1-\alpha)g(1,y')][\alpha f(2,y) + (1-\alpha)g(2,y)] \\
& \quad = [\alpha f(1,y) + (1-\alpha)g(1,y)][\alpha \left(1 - \frac{1}{2} f(1,y')\right) + (1-\alpha)\left(1 - \frac{1}{2} g(1,y')\right)] \\
& \quad\quad -[\alpha f(1,y') + (1-\alpha)g(1,y')][\alpha \left(1 - \frac{1}{2} f(1,y)\right) + (1-\alpha)\left(1 - \frac{1}{2} g(1,y)\right)] \\
& \quad = [\alpha f(1,y) + (1-\alpha)g(1,y)][1 - \frac{1}{2}\left(\alpha f(1,y') + (1-\alpha)g(1,y')\right)]
\end{aligned}$$

$$\begin{aligned}
&-[\alpha f(1,y') + (1-\alpha)g(1,y')][1 - \frac{1}{2}(\alpha f(1,y) + (1-\alpha)g(1,y))] \\
&= \alpha f(1,y) + (1-\alpha)g(1,y) - [\alpha f(1,y') + (1-\alpha)g(1,y')] \\
&= \alpha[f(1,y) - f(1,y')] + (1-\alpha)[g(1,y) - g(1,y')] \\
&\geq 0, \text{ by Proposition 2.}
\end{aligned}$$

For the compactness of $M_\nu(\mathrm{TP}_2)$, let $M_\nu = \{\mu : \mu_2 = \nu\}$. Since $\mu_1(\{1,2\}) = 1$ and $\mu(\{1,2\} \times A) = \mu_2(A)$, it is immediate that M_ν is compact. Thus, since $M_\nu(\mathrm{TP}_2) \subset M_\nu$, it suffices to show that $M_\nu(\mathrm{TP}_2)$ is closed. But, this is immediate from Block, Savits and Shaked's Remark (vii) (1982).

Extreme Points. Now we embark on determining the extreme points of the compact convex set $M_\nu(\mathrm{TP}_2)$ and obtain a representation of μ in $M_\nu(\mathrm{TP}_2)$ in terms of extreme points of $M_\nu(\mathrm{TP}_2)$. We are not entirely successful.

Let D be the support or spectrum of ν. D is the smallest closed subset of R with $\nu(D) = 1$. An equivalent description is: $x \in D$ if and only if $\nu\{(x-\epsilon, x+\epsilon)\} > 0$ for every $\epsilon > 0$.

For each $u \in D$, define μ_u on $\mathcal{B}$ by

$$\mu_u(\{1\} \times B_1 \cup \{2\} \times B_2) = \nu((-\infty, u] \cap B_1) + \nu((u, \infty) \cap B_2), \quad \text{for } B_1, B_2 \in \mathcal{C}.$$

It is easy to check that μ_u is a probability measure on $\mathcal{B}$ and $(\mu_u)_2 = \nu$. In an intuitive way, μ_u is built up on $\mathcal{B}$ by splitting ν into 2 parts: $\{1\} \times (-\infty, u]$ and $\{2\} \times (u, \infty)$. The compression of μ_u to a 2×2 table gives the following picture.

Description of μ_u			
$X \setminus Y$	$Y \leq u$	$Y > u$	Marginal sum
1	$\nu((-\infty, u])$	0	$\nu((-\infty, u]) = (\mu_u)_1(\{1\})$
2	0	$\nu((u, \infty))$	$\nu((u, \infty)) = (\mu_u)_1(\{2\})$
			1

Let the function $f_u : \{1,2\} \times R \to R$ be defined by

$$\begin{aligned}
f_u(1,y) &= 2 \quad \text{if} \quad -\infty < y \leq u, \\
&= 0 \quad \text{if} \quad u < y < \infty
\end{aligned}$$

and

$$\begin{aligned}
f_u(2,y) &= 0 \quad \text{if} \quad -\infty < y \leq u, \\
&= 2 \quad \text{if} \quad u < y < \infty.
\end{aligned}$$

It can be checked that f_u is a version of the Radon-Nikodym derivative of μ_u with respect to $\tau \otimes \nu$. From the description of f_u, it is clear that $\mu_u \in M_\nu(\mathrm{TP}_2)$.

Further, for distinct u_1 and u_2 in D, μ_{u_1} and μ_{u_2} are distinct. We now show that each μ_u is an extreme point of $M_\nu(\mathrm{TP}_2)$. Suppose $\mu_u = \alpha\mu + (1-\alpha)\lambda$ for some $\mu, \lambda \in M_\nu(\mathrm{TP}_2)$ and $0 \leq \alpha \leq 1$. Let f and g be versions of Radon-Nikodym derivatives of μ and λ, respectively, with respect to $\tau \otimes \nu$. Then

$$f_u(1,y) = \alpha f(1,y) + (1-\alpha)g(1,y) \text{ for almost all } y\ [\nu]$$

and

$$f_u(2,y) = \alpha f(2,y) + (1-\alpha)g(2,y) \text{ for almost all } y\ [\nu].$$

From Proposition 1 and the description of f_u, it follows that

$$f_u = f = g \text{ a.e. } [\tau \otimes \nu]$$

and

$$\mu_u = \mu = \lambda.$$

Now we come to the representation theorem. We need to distinguish several cases of D.

Case 1. D is bounded.

Let a and b be the left and right extremities of D, respectively. Note that a, $b \in D$. We distinguish two cases. Suppose a is an atom of ν, i.e., $\nu(\{a\}) > 0$. Define μ_{a^*} on $\mathcal{B}$ by

$$\mu_{a^*}(\{1\} \times B_1 \cup \{2\} \times B_2) = \nu(B_2) \text{ for all } B_1 \text{ and } B_2 \text{ in } \mathcal{C}.$$

The measure μ_{a^*} spreads the measure ν on line $x = 2$ leaving nothing for the line $x = 1$. Note that the probability measure μ_b spreads ν on the line $x = 1$ leaving nothing for the line $x = 2$. One can check that μ_{a^*} is an extreme point of $M_\nu(\mathrm{TP}_2)$ and distinct from μ_u for every u in D.

We conjecture that these are all the extreme points of $M_\nu(\mathrm{TP}_2)$. If this conjecture is true, then every measure μ in $M_\nu(\mathrm{TP}_2)$ is a mixture of extreme points of $M_\nu(\mathrm{TP}_2)$, i.e., there exists a probability measure λ on an appropriate σ-field on $D^* = \{a^*\} \cup D$ (which depends on μ) such that

$$\mu(A) = \int_{D^*} \mu_u(A)\lambda(du), \text{ for every } A \in \mathcal{B}. \tag{2}$$

The conjecture is true if D is finite. This can be shown as follows. Let $D = \{1, 2, \ldots, n\}$, $\mu \in M_\nu(\mathrm{TP}_2)$, $q_i = \nu(\{i\})$, $i = 1, 2, \ldots, n$, and $\mu(\{i,j\}) = p_{ij}$, $i = 1, 2$, $j = 1, 2, \ldots, n$. The Radon-Nikodym derivative of μ with respect to $\tau \otimes \nu$ works out to be

$$\begin{aligned} f(1,i) &= 2p_{1i}/q_i, \quad i = 1, 2, \ldots, n, \\ f(2,i) &= 2p_{2i}/q_i, \quad i = 1, 2, \ldots, n. \end{aligned}$$

The measure λ on $D^* = \{a^*\} \cup \{1, 2, \ldots, n\}$ is given by

$$\begin{aligned}
\lambda(\{a^*\}) &= 1 - p_{11}/q_1, \\
\lambda(\{1\}) &= p_{11}/q_1 - p_{12}/q_2, \\
\lambda(\{2\}) &= p_{12}/q_2 - p_{13}/q_3, \ldots, \\
\lambda(\{n-1\}) &= p_{1n-1}/q_{n-1} - p_{1n}/q_n, \\
\lambda(\{n\}) &= p_{1n}/q_n.
\end{aligned}$$

The representation (2) is then valid. See Subramanyam and Bhaskara Rao (1988).

The other possibility under Case 1 is that a is not an atom of ν. In this case, μ_{a^*} and μ_a are identical. There is no need to introduce μ_{a^*}. We again conjecture that the set of extreme points of $M_\nu(\mathrm{TP}_2)$ is precisely $\{\mu_u : u \in D\}$.

Case 2. D is unbounded.

Assume that D is unbounded on both sides. Introduce two new measures $\mu_{-\infty}$ and μ_∞ by

$$\mu_{-\infty}(\{1\} \times B_1 \cup \{2\} \times B_2) = \nu(B_1)$$

and

$$\mu_\infty(\{1\} \times B_1 \cup \{2\} \times B_2) = \nu(B_2) \text{ for all } B_1, B_2 \in \mathcal{C}.$$

Then $\mu_{-\infty}$, $\mu_\infty \in M$ (TP_2), $\mu_{-\infty}$ is concentrated on the line $x = 1$, μ_∞ on the line $x = 2$, and $\mu_{-\infty}$ and μ_∞ are extreme points of $M_\nu(\mathrm{TP}_2)$. We again conjecture that $\mu_{-\infty}$, μ_∞, μ_u, $u \in D$ are the only extreme points of $M_\nu(\mathrm{TP}_2)$.

The last case that D is unbounded on one side only can be discussed in a similar vein.

REFERENCES

BARLOW, R.E. and PROSCHAN, F. (1981). *Statistical Theory of Reliability and Life Testing.* To Begin With, Silver Spring, MD.

BLOCK, H.W., SAVITS, T.H., and SHAKED, M. (1982). Some concepts of negative dependence. *Ann. Prob.* **3** 765–772.

HALMOS, P.R. (1950). *Measure Theory.* Van Nostrand, New York.

KARLIN, S. (1968). *Total Positivity.* Stanford University Press, Stanford.

LEHMANN, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **43** 1137–1153.

SUBRAMANYAM, K. and BHASKARA RAO, M. (1988). Analysis of odds ratios in $2 \times n$ ordinal contingency tables. *J. Multiv. Anal.* **27** 478–493.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF NORTH CAROLINA
WILMINGTON, NC 28403

DEPARTMENT OF STATISTICSS
NORTH DAKOTA STATE UNIVERSITY
FARGO, ND 58105