Article

# The Ordering of Shannon Entropies for the Multivariate Distributions and Distributions of Eigenvalues 

Ming-Tien Tsai ${ }^{1, *}$, Feng-Ju Hsu ${ }^{2}$ and Chia-Hsuan Tsai ${ }^{3}$<br>1 Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan<br>2 Center for General Education, Saint Mary's Junior College of Medicine, Nursing and Management, Yi-Lang 266, Taiwan; eddahsu@gmail.com<br>3 Institute of Statistics, National Chiao-Tung University, Hsin-Chu 30010, Taiwan; teresanctu.st02g@nctu.edu.tw<br>* Correspondence: mttsai@stat.sinica.edu.tw

Received: 14 January 2019; Accepted: 18 February 2019; Published: 20 February 2019


#### Abstract

In this paper, we prove the Shannon entropy inequalities for the multivariate distributions via the notion of convex ordering of two multivariate distributions. We further characterize the multivariate totally positive of order $2\left(M T P_{2}\right)$ property of the distribution functions of eigenvalues of both central Wishart and central MANOVA models, and of both noncentral Wishart and noncentral MANOVA models under the general population covariance matrix set-up. These results can be directly applied to both the comparisons of two Shannon entropy measures and the power monotonicity problem for the MANOVA problem.


Keywords: central Wishart and central MANOVA models; $M T P_{2}$; noncentral Wishart and noncentral MANOVA models

## 1. Introduction

Let $\mathbf{X}_{1}, \mathbf{X}_{2}, \cdots$ be independent random vectors with a common continuous distribution function $F(\mathbf{x})$ which has density function $f(\mathbf{x}), \mathbf{x} \in R^{p}$ with respect to Lebesgue measure. The Shannon entropy for $f(\mathbf{x})$ is denoted by

$$
\begin{equation*}
\mathrm{I}_{S}(f)=\int_{-\infty}^{\infty}\{-\log f(\mathbf{x})\} f(\mathbf{x}) d \mathbf{x} \tag{1}
\end{equation*}
$$

We further use $\mathcal{F}$ to denote all density functions whose Shannon entropies exist. That is

$$
\begin{equation*}
\mathcal{F}=\left\{f: \mathrm{I}_{S}(f)<\infty\right\} \tag{2}
\end{equation*}
$$

Entropy inequality has been studied by several authors, such as Karlin and Rinott ([1,2], and references therein), who used the notions of majorization and Schur functions to study the Shannon (differential) entropy comparisons. Zografos and Nadarajah [3] pointed out that some comparisons of Shannon entropies can be induced by the affine transformation, those results show agreement on the concern that Shannon entropy is not scale invariant. Tsai [4] used the notion of convex ordering to investigate the Shannon inequalities for the univariate distributions. The novel finding of his approach is to adopt the difference of two Shannon entropy measures as a new measure, which is essentially symmetric. This new measure has some advantages over the well-known Kullback-Leibler divergence (relative entropy). The former one is designed for small or moderate deviation, however, the latter one is quite often used for large deviation. The measure of the difference between two Shannon entropies enjoys the finite sample property.

The key point of this method is to transform the convex ordering into the Lorenz ordering, which nicely depicts the stochastic ordering of two normalized distributions with the same starting and ending points, respectively. The presentation of the difference between two Shannon entropy measures can be represented in terms of a standard form of Kullback-Leibler divergence, however, it is interesting to note that the original Kullback-Leibler divergence does not have such kind of symmetric representation. Furthermore, the new measure is monotone likelihood ratio of two normalized density functions. This new measure provides us with some advantages for statistical inference.

The ordering of Shannon entropies can naturally occur when two underline distributions have the convex ordering or concave ordering relationship via the help of Lorenz ordering. In this paper, one of the main goals is to extend the result of Tsai [4] to the multivariate case, which is studied in Section 3. Please note that the notions of convex ordering of two distributions, monotone likelihood ratio $(M L R)$ and totally positive of order $2\left(T P_{2}\right)$ are essentially equivalent. We then adopt the notion of multivariate totally positive of order $2\left(M T P_{2}\right)$ or multivariate reverse rule of order $2\left(M R R_{2}\right)$ to get a more complete theory on Shannon entropy comparisons. Eigenvalues problem has become quite a hot topic at present. We further study the $M T P_{2}$ property of the distribution functions of eigenvalues of both central Wishart and central MANOVA models, and of both noncentral Wishart and noncentral MANOVA models under the general population covariance matrix set-up, respectively, in Section 4. The results can be directly applied to the comparisons of two Shannon entropy measures as well as the power monotonicity problem for the MANOVA problem, which has been open for a long time in the literature (Perlman and Olkin [5]). For a high-dimensional Wishart matrix with unknown scale times identity matrix as the population matrix, the $T P_{2}$ property of limiting empirical density function of eigenvalues is also studied in the Section 4.5. Mixture density functions, such as the noncentral chi-square density function and those presented in Sections 4.3 and 4.4, play an important role in statistical inference. We also connected this notion with that of well-known Fisher's multivariate analysis of variance in the final remark section.

## 2. The Univariate Distributions

Tsai [4] adopted the notion of convex ordering of distribution functions to study the orderings of Shannon entropy measures. For two univariate distributions $F, G \in \mathcal{F}$, it is said that $F$ is $c$-ordered (convex ordered) with respect to $G\left(F \leq_{c} G\right)$ if and only if $G^{-1} F$ is convex on the interval where $0<F(x)<1$ (van Zwet [6]). The transformation of Barlow and Doksum [7] is to transform the convex ordering of distributions $F$ and $G$ to the stochastic ordering of distributions, it essentially can be viewed as a kind of Lorenz ordering (Gastwirth [8]).

Let $C_{F}(x)=G^{-1} F(x)$, then note that $d C_{F}(x) / d x=f(x) / g\left(G^{-1} F(x)\right), x \in(-\infty, \infty)$. It can also be rewritten as $f\left(F^{-1}(u)\right) / g\left(G^{-1}(u)\right)$, where $u=F(x), u \in[0,1]$. We adopt the notion of Barlow and Doksum [7] to denote that $H_{F}(u)=\int_{0}^{G^{-1}(u)} f\left(F^{-1} G(t)\right) d t, u \in[0,1]$, which can be viewed as the inverse function of the transformation considered by Barlow and Doksum [1]. Please note that the notation in Barlow and Doksum [7] is originally defined that $F(0)=0$ and $F^{-1}(0)=0$, we use the general notations that $F(-\infty)=0$ and $F^{-1}(0)=-\infty$ in this paper. Obviously, $H_{G}(u)=u, u \in[0,1]$. Under the assumption that $F(x) \leq_{c} G(y)$, Barlow and Doksum [7] pointed out that both $H_{F}(u)$ and $H_{G}(u), u \in[0,1]$ are two distribution functions. Let $h_{F}$ and $h_{G}$ be the corresponding density functions of $H_{F}$ and $H_{G}$, respectively, then we may see that $h_{F}(u)=f\left(F^{-1}(u)\right) / g\left(G^{-1}(u)\right)$ and $h_{G}(u)=1, u \in[0,1]$. Then, it is easy to note that the difference of two Shannon entropies can be represented as the Kullback-Leibler divergence of the uniform density function $h_{G}$ and the density function $h_{F}$.

Please note that if $F(x) \leq_{c} G(y)$, then $h_{F}(u)$ is nondecreasing in $u \in[0,1]$, namely, it enjoys the property of monotone likelihood ratio. In addition, it gets across with the uniform density function $h_{G}(u)=1, u \in[0,1]$ at most once, the sign of the difference $\left(h_{F}(u)-h_{G}(u), u \in[0,1]\right)$ of two density functions is from negative to nonnegative. Also note that $\int_{0}^{1}\left(h_{F}(v)-1\right) d v=0$, thus we have that $H_{F}(u) \leq u=H_{G}(u), u \in[0,1]$. Similarly, if $F(x) \leq_{c} G(y)$, then $F(x) \leq G(y)$, and $C_{F}(x)$ is a convex function in $x, x \in(-\infty, \infty)$ (for the details see Shaked and Shanthikumar [9]). Namely, the convex
ordering implies the stochastic ordering, however, the vice versa may not be true. We provide a simple counterexample for this. For $0 \leq x \leq 1$, let $F(x)=x, G(x)=(3 x-2)^{3} / 9+8 / 9$, then $F^{-1}(u)=u, G^{-1}(u)=(\sqrt[3]{9 u-8}+2) / 3, u \in[0,1]$. Thus, $f\left(F^{-1}(u)\right)=1, g\left(G^{-1}(u)\right)=(9 u-8)^{2 / 3}$ and $h_{F}(u)=(9 u-8)^{2 / 3}$. It is easy to see that $F(u) \leq G(u)$. However, $h_{F}(u)$ is not nondecreasing in $u$, $u \in[0,1]$, that means that $F(x) \leq_{c} G(y)$ will not hold in this example. As such, we may conclude that if $F(x) \leq_{c} G(y)$, then $F(x) \leq G(y)$. However, if $F(x) \leq G(y)$, then $F(x) \leq_{c} G(y)$ may not hold generally.

Theorem 1 (Tsai [4]). If $F(x) \leq_{c} G(y)$, then $I_{S}(f) \geq I_{S}(g)$.
With the convex ordering assumption, i.e., $F(x) \leq_{c} G(y)$, the corresponding Lorenz ordering plays an important role for the proof of Theorem 1. To check whether the condition $F(x) \leq_{c} G(y)$ holds or not, one should make simultaneously check whether the condition $F(x) \leq G(y)$ holds or not. Similarly, define that $F(x) \leq_{\text {concave }} G(x)$ if and only if $F^{-1} G(x)$ is concave function in $x, x \in(-\infty, \infty)$, then we have the following corollary.

Corollary 1. If $F(x) \leq_{\text {concave }} G(y)$, then $I_{S}(f) \leq I_{S}(g)$.
Proof. It can be easily followed by Theorem 1 via changing the sign of $\log h_{F}(u)$.
Proposition 1. Composition lemma (Karlin [10]). If $K$ is $T P_{2}$ and $L$ is $T P_{2}$ and $v$ is a $\sigma$-finite measure, then the convolution $M(x, y)=\int_{Z} K(x, z) L(z, y) d v(z)$ is $T P_{2}$.

It can be proved similarly. We also note that if $K$ is $T P_{2}$ and $L$ is $R R_{2}$, then $M$ is $R R_{2}$.

## 3. The Multivariate Distributions

By convex ordering transformation, we can neatly transform the difference of two Shannon entropy measures into the Lorenz ordering. In this section, we intend to extend this result to the multivariate case. It is natural to do so in the sense of conditional convex ordering for the multivariate case. Let $\mathbf{X}=\left(X_{1}, \ldots, X_{p}\right)^{\top}$ and $\mathbf{Y}=\left(Y_{1}, \ldots, Y_{p}\right)^{\top}$ be random vectors, and $f$ and $g$ be the density function of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Please note that the $M T P_{2}$ holds if and only if all the pairwise $T P_{2}$ hold (Fact 4.3.2 of Tong [11]), namely to examine the $M T P_{2}$ is the same as to examine all the pairwise $T P_{2}$ for any fixed pair. Hence without loss of generality, we may take $p=2$. Please note that $f\left(x_{1}, x_{2}\right)=f_{1}\left(x_{1}\right) f_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$, thus $f\left(x_{1}, x_{2}\right)$ is $T P_{2}$ if and only if $f_{1}\left(x_{1}\right) f_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$ is $T P_{2}$. By the product property that the product function is $T P_{2}$ if two functions are each $T P_{2}$, respectively. Thus we may conclude that $f\left(x_{1}, x_{2}\right)$ is $T P_{2}$ if and only if both $f_{1}\left(x_{1}\right)$ and $f_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$ are $T P_{2}$. That is to say that $F\left(x_{1}, x_{2}\right) \leq_{c} G\left(y_{1}, y_{2}\right)$ iff $F_{1}\left(x_{1}\right) \leq_{c} G_{1}\left(y_{1}\right)$ and $F_{2 \mid 1}\left(x_{2} \mid x_{1}\right) \leq_{c} G_{2 \mid 1}\left(y_{2} \mid y_{1}\right)$. Continue this process, for the general $p$ case, we have the following.

Proposition 2. Let $\mathbf{c}=\left(c_{1}, \ldots, c_{p}\right)^{\top} \in R^{p}$, then (i) $P\left\{Y_{1}>c_{1}\right\} \leq_{c} P\left\{X_{1}>c_{1}\right\}$, and (ii) $P\left\{Y_{j}>c_{j} \mid Y_{1}=\right.$ $\left.y_{1}, \ldots, Y_{j-1}=y_{j-1}\right\} \leq_{c} P\left\{X_{j}>c_{j} \mid X_{1}=x_{1}, \ldots, X_{j-1}=x_{j-1}\right\}$ for all $x_{1} \leq y_{1}, \ldots, x_{j-1} \leq y_{j-1}$ and for all $j=2, \ldots, p$ if and only if $F(\mathbf{x}) \leq_{c} G(\mathbf{y})$.

From the results of Proposition 2, we are now in a position to use the convex ordering assumption for which $F_{j \mid 1, \ldots, j-1}\left(x_{j} \mid x_{1}, \ldots, x_{j-1}\right) \leq_{c} G_{j \mid 1, \ldots, j-1}\left(y_{j} \mid y_{1}, \ldots, y_{j-1}\right)$ to get each corresponding Lorenz ordering, $\forall j=1, \ldots, p$, where $j=1$ denotes that $F_{1}\left(x_{1}\right) \leq_{c} G_{1}\left(y_{1}\right)$. With some elementary arguments, we then have the following.

Theorem 2. If $F(\mathbf{x}) \leq_{c} G(\mathbf{y})$, then $I_{S}(f) \geq I_{S}(g)$.
Proof. For simplicity we may consider the situation when $p=2$ first. Let $g_{12}\left(y_{1}, y_{2}\right)$ be the joint density function of random variables $Y_{1}$ and $Y_{2}, g_{2 \mid 1}\left(y_{2} \mid y_{1}\right)$ be the corresponding conditional density function of $Y_{2}$ given $Y_{1}=y_{1}$ and $G_{2 \mid 1}\left(y_{2} \mid y_{1}\right)$ be the corresponding distribution function.

Let $f_{12}\left(x_{1}, x_{2}\right), f_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$ and $F_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$ be defined similarly for the variables $X_{1}$ and $X_{2}$. Also let $U_{1}=F_{1}(x), U_{2} \mid u_{1}=F_{2 \mid 1}\left(x_{2} \mid x_{1}\right)\left(V_{1}=G_{1}(y), V_{2} \mid v_{1}=G_{2 \mid 1}\left(y_{2} \mid y_{1}\right)\right)$. Please note that

$$
\begin{align*}
\mathrm{I}_{S}\left(f_{12}\right)-\mathrm{I}_{S}\left(g_{12}\right)= & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty}\left\{-\log f_{12}\left(x_{1}, x_{2}\right)\right\} f_{12}\left(x_{1}, x_{2}\right) d x_{1} d x_{2}  \tag{3}\\
& -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty}\left\{-\log g_{12}\left(y_{1}, y_{2}\right)\right\} g_{12}\left(y_{1}, y_{2}\right) d y_{1} d y_{2} \\
= & \int_{0}^{1} \int_{0}^{1}\left\{-\log \left[f_{1}\left(F_{1}^{-1}\left(u_{1}\right)\right) f_{2 \mid 1}\left(F_{2 \mid 1}^{-1}\left(u_{2} \mid u_{1}\right)\right)\right]\right\} d u_{1} d u_{2} \\
& -\int_{0}^{1} \int_{0}^{1}\left\{-\log \left[g_{1}\left(G_{1}^{-1}\left(v_{1}\right)\right) g_{2 \mid 1}\left(G_{2 \mid 1}^{-1}\left(v_{2} \mid v_{1}\right)\right)\right]\right\} d v_{1} d v_{2} \\
= & \int_{0}^{1} \int_{0}^{1}-\log \left[\frac{f_{1}\left(F_{1}^{-1}\left(u_{1}\right)\right) f_{2 \mid 1}\left(F_{2| |}^{-1}\left(u_{2} \mid u_{1}\right)\right)}{g_{1}\left(G_{1}^{-1}\left(u_{1}\right)\right) g_{2 \mid 1}\left(G_{2 \mid 1}^{-1}\left(u_{2} \mid u_{1}\right)\right)}\right] d u_{1} d u_{2} \\
= & \int_{0}^{1}-\log h_{F_{1}}\left(u_{1}\right) d u_{1}+\int_{0}^{1}\left[\int_{0}^{1}-\log h_{F_{2 \mid 1}}\left(u_{2} \mid u_{1}\right) d u_{2}\right] d u_{1}
\end{align*}
$$

where $h_{F_{1}}\left(u_{1}\right)=f_{1}\left(F_{1}^{-1}\left(u_{1}\right)\right) / g_{1}\left(G_{1}^{-1}\left(u_{1}\right)\right), h_{G_{1}}\left(u_{1}\right)=1$ and $h_{F_{2 \mid 1}}\left(u_{2} \mid u_{1}\right)=f_{2 \mid 1}\left(F_{2 \mid 1}^{-1}\left(u_{2} \mid u_{1}\right)\right)$ $/ g_{2 \mid 1}\left(G_{2 \mid 1}^{-1}\left(u_{2} \mid u_{1}\right)\right), h_{G_{2 \mid 1}}\left(u_{2} \mid u_{1}\right)=1$. Furthermore, by Proposition 2 , we may note that $F\left(x_{1}\right) \leq_{c} G\left(y_{1}\right)$, and thus via the arguments in Section 2, we have that $h_{F_{1}}\left(u_{1}\right), u_{1} \in[0,1]$ is the density function. Similarly, we also have that $F_{2 \mid 1}\left(x_{2} \mid x_{1}\right) \leq_{c} G_{2 \mid 1}\left(y_{2} \mid y_{1}\right)$, and then $h_{F_{2 \mid 1}}\left(u_{2} \mid u_{1}\right), u_{2} \in[0,1]$ is the density function. Thus by the information inequality, the first term of the right hand side of Equation (3) is non-negative and the second term insight bracket is non-negative too, and hence the second term of the right hand side of Equation (3) is non-negative. Thus we may conclude that $\mathrm{I}_{S}\left(f_{12}\right) \geq \mathrm{I}_{S}\left(g_{12}\right)$.

Let $U_{i}\left|u_{1}, \ldots, u_{i-1}=F_{i \mid 1, \ldots, i-1}\left(x_{i} \mid x_{1}, \ldots, x_{i-1}\right), V_{i}\right| v_{1}, \ldots, v_{i-1}=G_{i \mid 1, \ldots, i-1}\left(y_{i} \mid y_{1}, \ldots, i_{p-1}\right)$ and $h_{F_{i \mid 1, \ldots, i-1}}\left(u_{i} \mid u_{1}, \ldots, u_{i-1}\right)=f_{i \mid 1, \ldots, i-1}\left(F_{i \mid 1, \ldots, i-1}^{-1}\left(u_{i} \mid u_{1}, \ldots, u_{i-1}\right)\right) / g_{i \mid 1, \ldots, i-1}\left(G_{i \mid 1, \ldots, i-1}^{-1}\right.$ $\left.\left(v_{i} \mid v_{1}, \ldots, v_{i-1}\right)\right), i=2, \ldots, p$. Continuing the processes, finally for general $p$ we have

$$
\begin{align*}
\mathrm{I}_{S}(f)-\mathrm{I}_{S}(g)= & \int_{-\infty}^{\infty}\{-\log f(\mathbf{x})\} f(\mathbf{x}) d \mathbf{x}-\int_{-\infty}^{\infty}\{-\log g(\mathbf{y})\} g(\mathbf{y}) d \mathbf{y}  \tag{4}\\
= & \int_{0}^{1}-\log h_{F_{1}}\left(u_{1}\right) d u_{1}+\int_{0}^{1}\left[\int_{0}^{1}-\log h_{F_{2 \mid 1}}\left(u_{2} \mid u_{1}\right) d u_{2}\right] d u_{1} \\
& +\ldots+\int_{0}^{1} \ldots \int_{0}^{1}\left[\int_{0}^{1}-\log h_{F_{p \mid 1, \ldots, p-1}}\left(u_{p} \mid u_{1}, \ldots, u_{p-1}\right) d u_{p}\right] d u_{p-1} \ldots d u_{1}
\end{align*}
$$

Hence, the theorem follows.
For two arbitrary densities $f_{g}, f_{h}$, the Kullback-Leibler information discrimination is defined by $K\left(f_{g}, f_{h}\right)=\int_{-\infty}^{\infty} f_{g}(x) \log \left\{f_{g}(x) / f_{h}(x)\right\} d x$. It is non-negative for all $f_{g}, f_{h}$, and is equal to zero if $f_{g}=f_{h}$ almost everywhere. Let $\mathcal{E}[X]$ denote the expectation of random variable $X$. If $F(\mathbf{x}) \leq_{c} G(\mathbf{y})$, from Equation (4) we have the following representation:

$$
\begin{equation*}
\mathrm{I}_{S}(f)-\mathrm{I}_{S}(g)=\mathrm{K}\left(1, h_{F_{1}}\right)+\mathcal{E}_{U_{1}}\left[\mathrm{~K}\left(1, h_{F_{2 \mid 1}}\right)\right]+\ldots+\mathcal{E}_{U_{1}, \ldots, U_{p-1}}\left[\mathrm{~K}\left(1, h_{F_{p \mid 1, \ldots, p-1}}\right)\right] \tag{5}
\end{equation*}
$$

where $\mathcal{E}_{U_{1}, \ldots, U_{i-1}}$ denotes the expectation of independent uniform densities of $\left(U_{1}, \ldots, U_{i-1}\right)$, $U_{i} \in[0,1], i=1, \ldots, p-1$.

Corollary 2. If $F(\mathbf{x}) \leq_{\text {concave }} G(\mathbf{y})$, then $I_{S}(f) \leq I_{S}(g)$.
As we have seen, the proof of Theorem 2 is surprisingly easy. However, to characterize the property of convex ordering for two multivariate distributions is sometimes much more complicated via the conditional approach than to characterize the $M T P_{2}$ property directly. The convex ordering
of two multivariate distributions is essentially equivalent to the notion of $M T P_{2}$, so does the concave ordering to the $M R R_{2}$. Karlin and Rinott $([12,13])$ had studied several $M T P_{2}$ and $M R R_{2}$ distributions, respectively. Some other often seen models related to Theorem 2 and Corollary 2 are exampled in the following.

Example 1. Let $\mathbf{y}=\mathbf{A x}$, where $\mathbf{A}$ is a nonsingular matrix. Then $I_{S}(g)=\log |\mathbf{A}|+I_{S}(f)$, where $|\mathbf{A}|$ denotes the determinant of matrix $\mathbf{A}$ (Zografos and Nadarajah [3]). Thus we have $I_{S}(f) \geq I_{S}(g)$ iff $|\mathbf{A}| \leq 1$.

Example 2. Let $\mathcal{B}^{p}=\left\{\left(x_{1}, \ldots, x_{p}\right)^{\top}: x_{i}>0, i=1, \ldots, p, \sum_{i=1}^{p} x_{i} \leq 1\right\}$ be the $p$-dimensional simplex, and let

$$
f\left(\mathbf{x}, a_{1}, \ldots, a_{p} ; b_{1}, \ldots, b_{p}\right)=c \prod_{i=1}^{p} x_{i}^{a_{i}-1}\left(1-\sum_{k=1}^{i} x_{k}\right)^{b_{i}-1}
$$

be the generalized Dirichlet density function with parameters, $a_{i}>0, b_{i}>0, i=1, \ldots, p$ and

$$
c=\prod_{i=1}^{p} \frac{\Gamma\left(\sum_{k=i}^{p}\left(a_{k}+b_{k}-1\right)+1\right)}{\Gamma\left(a_{i}\right) \Gamma\left(b_{i}+\sum_{k=i}^{p}\left(a_{k}+b_{k}-1\right)\right)} .
$$

It is easy to note that $x_{i}^{a_{i}-1}$ is $T P_{2}$ in $\left(x_{i}, a_{i}\right),\left(1-\sum_{k=1}^{i} x_{k}\right)^{b_{i}-1}$ is $R R_{2}$ in $\left(x_{i}, b_{i}\right)$ and $\left(1-\sum_{k=1}^{i} x_{k}\right)^{b_{i}-1}$ is pairwisely $R R_{2}$ in $\left(x_{i}, x_{j}\right)$ when $b_{i} \geq 1$. Then, by Proposition 2 , we have (i) $\prod_{i=1}^{p} x_{i}^{a_{i}-1}$ is $M T P_{2}$ and (ii) $\prod_{i=1}^{p}\left(1-\sum_{k=1}^{i} x_{k}\right)^{b_{i}-1}$ is $M R R_{2}$ when $b_{i} \geq 1, \forall i=1, \ldots, p$. Thus, we may conclude that the generalized Dirichlet density function is $M R R_{2}$ when $b_{i} \geq 1, \forall i=1, \ldots, p$. In addition, Corollary 2 is applicable.

Example 3. Gupta and Richards [14] considered the multivariate Liouville distribution which the density function is of the form

$$
f(\mathbf{x}, \boldsymbol{\theta})=c(\boldsymbol{\theta}) g\left(\sum_{i=1}^{p} x_{i}\right) \prod_{i=1}^{p} x_{i}^{\theta_{i}}
$$

where $g: R^{+} \rightarrow R^{+}$is continuous, and $\boldsymbol{\theta}, \mathbf{x} \in R^{+p}$ and

$$
c^{-1}(\boldsymbol{\theta})=\frac{\prod_{i=1}^{p} \Gamma\left(\theta_{i}\right)}{\Gamma\left(\sum_{i=1}^{p} \theta_{i}\right)} \int_{0}^{\infty} t^{\sum_{i=1}^{p} \theta_{i}-1} g(t) d t
$$

Please note that $\prod_{i=1}^{p} x_{i}^{\theta_{i}}$ is $M T P_{2}$. By Proposition 2, it is easy to see that $f(\mathbf{x}, \boldsymbol{\theta})$ is $M T P_{2}\left(M R R_{2}\right)$ if and only if $g\left(x_{1}+x_{2}\right)$ is $T P_{2}\left(R R_{2}\right)$ in $\left(x_{1}, x_{2}\right)$ on $R^{+2}$. The $M R R_{2}$ property of some special cases of multivariate Liouville distributions, such as Dirichlet distributions and inverted Dirichlet distributions, are studied by Karlin and Rinott [13].

Example 4. In biostatistics, many studies are under the setup of well-known Cox proportional hazard model (Cox [15]), and its basic model is assumed to be that $\bar{G}(\mathbf{x})=(\bar{F}(\mathbf{x}))^{\lambda}, \lambda>0$, where $\bar{G}(\mathbf{x})=$ $1-G(\mathbf{x})$. Without loss of generality, we may assume $p=2$. Write $\bar{G}(\mathbf{x})=\bar{G}_{1}\left(x_{1}\right) \bar{G}_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$ and $\bar{F}(\mathbf{x})=\bar{F}_{1}\left(x_{1}\right) \bar{F}_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$. Let $G_{1}(x)=u_{1}, u_{1} \in[0,1]$, then we have $F_{1}\left(G_{1}^{-1}\left(u_{1}\right)\right)=1-\left(1-u_{1}\right)^{1 / \lambda}$, thus we have $d F_{1}\left(G_{1}^{-1}\left(u_{1}\right)\right) / d u_{1}=\left(1-u_{1}\right)^{(1 / \lambda-1)} / \lambda$, which is nondecreasing in $u_{1}$ if $\lambda \geq 1$. As such, we have $G_{1}^{-1}\left(u_{1}\right) \leq_{c} F_{1}^{-1}\left(u_{1}\right), u_{1} \in[0,1]$ when $\lambda \geq 1$. Namely, when $\lambda \geq 1$ we have $F_{1}\left(x_{1}\right) \leq_{c} G_{1}\left(x_{1}\right)$. Let $G_{2 \mid 1}\left(x_{2} \mid x_{1}\right)=u_{2}$, similar arguments as above, we may have that $F_{2 \mid 1}\left(x_{2} \mid x_{1}\right) \leq_{c} G_{2 \mid 1}\left(x_{2} \mid x_{1}\right)$ when $\lambda \geq 1$. Thus, by Proposition 2 , we may have that $F(\mathbf{x}) \leq_{c} G(\mathbf{x})$ when $\lambda \geq 1$. This procedure can be directly generalized to the general $p$ case by mathematical induction. As such, for the positive disease gene case $\lambda \leq 1$, then by Theorem 2 we may conclude that $I_{S}(f) \leq I_{S}(g)$. Similarly, for the negative disease gene case $\lambda \geq 1$, then $I_{S}(f) \geq I_{S}(g)$.

Example 5. The density function of eigenvalues for a multivariate beta matrix is of the form

$$
f(\mathbf{x})=c \prod_{i=1}^{p} x_{i}^{\frac{1}{2}\left(n_{1}-p+1\right)-1}\left(1-x_{i}\right)^{\frac{1}{2}\left(n_{2}-p+1\right)-1} \prod_{1 \leq i<j \leq p}\left|x_{i}-x_{j}\right|
$$

where $0<x_{i}<1, i=1, \ldots, p, n_{1}, n_{2}>p$, and

$$
c=\prod_{i=1}^{p} \frac{\Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{1}{2}\left(n_{1}+n_{2}-p+i\right)\right)}{\Gamma\left(1+\frac{i}{2}\right) \Gamma\left(\frac{1}{2}\left(n_{1}-p+i\right)\right) \Gamma\left(\frac{1}{2}\left(n_{1}-p+i\right)\right)}
$$

(Dumitriu [16]; Peddada and Richards [17]). Please note that the Vandermonde determinant $\prod_{1 \leq i<j \leq p}\left|x_{i}-x_{j}\right|$ is $M_{2}$, for the details see Dykstra and Hewett [18]. Using Proposition 2, it is easy to see that $\prod_{i=1}^{p} x_{i}^{\left(n_{1}-p+1\right) / 2-1}$ is $T P_{2}$ in $\left(x_{i}, n_{1}-p+1\right)$, and $\prod_{i=1}^{p}\left(1-x_{i}\right)^{\left(n_{2}-p+1\right) / 2-1}$ is $R R_{2}$ in $\left(x_{i}, n_{2}-p+1\right)$. Thus, $\prod_{i=1}^{p} x_{i}^{\left(n_{1}-p+1\right) / 2-1}\left(1-x_{i}\right)^{\left(n_{2}-p+1\right) / 2-1}$ is $M R R_{2}$. Then, we may conclude that the density function, which is the product of these functions, is $M R R_{2}$. In addition, Corollary 2 is applicable.

## 4. The Central Wishart and Central MANOVA Models, and Noncentral Wishart and Noncentral MANOVA Models

Theorem 2 states that if the density function has the $M T P_{2}$ property, then the corresponding ordering of Shannon entropies holds, namely the $M T P_{2}$ (or $M R R_{2}$ ) property of density function ensures the ordering of Shannon entropies. The $M T P_{2}$ also implies the power monotonicity property of MANOVA tests, for which based on the monotone function of eigenvalues. Eigenvalues play an important role in statistical inference. Example 5 deals with the density function of eigenvalues of MANOVA models (or Jacobi ensembles) when the population covariance matrix is assumed to be an identity matrix $(\Sigma=\mathbf{I})$. Two other often seen models are (i) Gaussian (or Hermite) ensembles,

$$
f(\mathbf{x})=(2 \pi)^{-\frac{p}{2}} \prod_{i=1}^{p} \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(1+\frac{i}{2}\right)} e^{-\frac{1}{2} \sum_{1}^{p} x_{i}^{2}} \prod_{1 \leq i<j \leq p}\left|x_{i}-x_{j}\right|,
$$

and (ii) Wishart models (or Laguerre ensembles),

$$
f(\mathbf{x})=2^{-\frac{p(p+1)}{2}} \prod_{i=1}^{p} \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(1+\frac{i}{2}\right) \Gamma\left(\frac{1}{2}\left(n_{1}-p+1\right)\right)} \prod_{i=1}^{p} x_{i}^{\frac{1}{2}\left(n_{1}-p+1\right)-1} e^{-\frac{1}{2} \sum_{1}^{p} x_{i}} \prod_{1 \leq i<j \leq p}\left|x_{i}-x_{j}\right|
$$

(Dumitriu [16]). From similar arguments to Example 5, it is easy to see that the density function $f(\mathbf{x})$ is $M T P_{2}$ either for (i) (the identity population covariance matrix is an M-matrix) or for (ii) (gamma type density function), respectively, we omit the details.

For statistical inference, generally the population covariance matrix is unknown. In this section, we further study the $M T P_{2}$ property of the distribution functions of eigenvalues of central Wishart and MANOVA models, and of noncentral Wishart and MANOVA models (James [19]; Muirhead [20]) under the general population covariance matrix set-up.

### 4.1. Type ${ }_{0} F_{0}$, Exponential: Eigenvalues of a Wishart Matrix

Suppose that the columns of a matrix $p \times n$ matrix $\mathbf{X}$ are independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ and $\mathrm{E}(\mathbf{X})=\mathbf{0}$. Let $\mathbf{L}=\operatorname{diag}\left(l_{1}, \ldots, l_{p}\right)$, where $l_{i}$ be the $i$ th largest eigenvalue of $\mathbf{X} \mathbf{X}^{\top}$. Similarly, let $\boldsymbol{\Delta}=\operatorname{diag}\left(\delta_{1}, \ldots, \delta_{1}\right)$, where $\delta_{i}$ be the $i$ th largest eigenvalue of $\boldsymbol{\Sigma}$. Write $\boldsymbol{\Sigma}=\mathbf{Q} \boldsymbol{\Delta} \mathbf{Q}^{\top}$
and $\mathbf{X X} \mathbf{X}^{\top}=\mathbf{U L U}^{\prime}$, where $\mathbf{Q}, \mathbf{U} \in \mathcal{O}(p)$ with $\mathcal{O}(p)$ being the group of $p \times p$ orthogonal matrices. Also denote $(d \mathbf{U})$ the $m$-form of $\mathbf{U}$ with $m=\frac{1}{2} p(p+1)$ and tr denotes the trace operator, then

$$
\begin{align*}
{ }_{0} F_{0}\left(-\frac{1}{2} \boldsymbol{\Sigma}^{-1}, \mathbf{X} \mathbf{X}^{\top}\right) & =\int_{\mathcal{O}(p)} e^{-\frac{1}{2} \operatorname{tr}\left(\mathbf{Q} \mathbf{\Delta}^{-1} \mathbf{Q}^{\top} \mathbf{U L \mathbf { U } ^ { \top }}\right)}(d \mathbf{U})  \tag{6}\\
& =\int_{\mathcal{O}(p)} e^{-\frac{1}{2} \operatorname{tr}\left(\Delta^{-1} \mathbf{H L H}^{\top}\right)}(d \mathbf{H}) \\
& ={ }_{0} F_{0}\left(-\frac{1}{2} \boldsymbol{\Delta}^{-1}, \mathbf{L}\right)
\end{align*}
$$

where $\mathbf{H}=\mathbf{Q}^{\top} \mathbf{U} \in \mathcal{O}(p)$ and noting that $(d \mathbf{U})=(d \mathbf{H})$. Due to the invariant property, the density function of $\mathbf{L}$ depends only upon $\Delta$, which is of the form

$$
\begin{equation*}
\phi_{\delta}\left(l_{1}, \ldots, l_{p}\right)=|\Delta|^{-\frac{1}{2} n}{ }_{0} F_{0}\left(-\frac{1}{2} \boldsymbol{\Delta}^{-1}, \mathbf{L}\right) \phi_{0}\left(l_{1}, \ldots, l_{p}\right), \tag{7}
\end{equation*}
$$

where

$$
\begin{equation*}
\phi_{0}\left(l_{1}, \ldots, l_{p}\right)=\frac{\pi^{\frac{1}{2} p^{2}}}{2^{\frac{1}{2} p n} \Gamma_{p}\left(\frac{1}{2} n\right) \Gamma_{p}\left(\frac{1}{2} p\right)}|\mathbf{L}|^{\frac{1}{2}(n-p-1)} \prod_{i<j}\left(l_{i}-l_{j}\right) \tag{8}
\end{equation*}
$$

with $\Gamma_{p}(a)=\pi^{\frac{1}{4} p(p-1)} \sum_{i=1}^{p} \Gamma\left(a-\frac{1}{2}(i-1)\right)$.
Theorem 3. ${ }_{0} F_{0}\left(-\frac{1}{2} \Delta^{-1}, \mathbf{L}\right)$ is $M T P_{2}$.
Proof. Please note that

$$
\begin{aligned}
\frac{\partial \log _{0} F_{0}\left(-\frac{1}{2} \boldsymbol{\Delta}^{-1}, \mathbf{L}\right)}{\partial \mathbf{L}} & =\frac{\int_{\mathcal{O}(p)} e^{-\frac{1}{2} \operatorname{tr}\left(\Delta^{-1} \mathbf{H L H}^{\top}\right)}\left(-\frac{1}{2} \boldsymbol{\Delta}^{-1} \mathbf{H} \mathbf{H}^{\top}\right)(d \mathbf{H})}{\int_{\mathcal{O}(p)} e^{-\frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Delta}^{-1} \mathbf{H L H}^{\top}\right)}(d \mathbf{H})} \\
& =\frac{\int_{\mathcal{O}(p)} e^{-\frac{1}{2} \operatorname{tr}\left(\Delta^{-1} \mathbf{H L H}^{\top}\right)}\left(-\frac{1}{2} \Delta^{-1}\right)(d \mathbf{H})}{\int_{\mathcal{O}(p)} e^{-\frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Delta}^{-1} \mathbf{H L H}^{\top}\right)}(d \mathbf{H})} \\
& =-\frac{1}{2} \Delta^{-1} .
\end{aligned}
$$

Thus, it is easy to see that

$$
\frac{\partial^{2} \log _{0} F_{0}\left(-\frac{1}{2} \Delta^{-1}, \mathbf{L}\right)}{\partial \mathbf{L}^{\prime} \partial \mathbf{L}}=\mathbf{0}
$$

and

$$
\frac{\partial^{2} \log _{0} F_{0}\left(-\frac{1}{2} \Delta^{-1}, \mathbf{L}\right)}{\partial \Delta \partial \mathbf{L}}=\frac{1}{2} \Delta^{-2}
$$

respectively. Please note that $\Delta$ is a positive diagonal matrix, which is clearly positive definite. Hence the theorem follows.

By the result of Dykstra and Hewett [18], it is easy to note that $\phi_{0}\left(l_{1}, \ldots, l_{p}\right)$ enjoys the $M T P_{2}$ property. Hence, by Theorem 3 we may conclude that the density function $\phi_{\delta}\left(l_{1}, \ldots, l_{p}\right)$ of eigenvalues of $\mathbf{X} \mathbf{X}^{\top}$, which is the product of two $M T P_{2}$ functions (Karlin and Rinott [12]), is $M T P_{2}$.

### 4.2. Type ${ }_{1} F_{0}$, Binomial Series: Eigenvalues When $\boldsymbol{\Sigma}_{1} \neq \boldsymbol{\Sigma}_{2}$

Suppose that the $p \times n_{1}$ matrix variate $\mathbf{X}$ with $n_{1}$ independent columns distributed as $N\left(\mathbf{0}, \boldsymbol{\Sigma}_{1}\right)$ and $p \times n_{2}$ matrix variate $\mathbf{Y}$ with $n_{2}$ independent columns distributed as $N\left(\mathbf{0}, \boldsymbol{\Sigma}_{2}\right)$. Also let $\mathbf{F}=$ $\operatorname{diag}\left(f_{1}, \ldots, f_{p}\right)$ and $\boldsymbol{\Omega}=\operatorname{diag}\left(\omega_{1}, \ldots, \omega_{p}\right)$, where $f_{i}$ is the $i$ th largest eigenvalue of

$$
\begin{equation*}
\left|\mathbf{X} \mathbf{X}^{\top}-f \mathbf{Y} \mathbf{Y}^{\top}\right|=0 \tag{9}
\end{equation*}
$$

and

$$
\begin{equation*}
\left|\Sigma_{1}-\omega \Sigma_{2}\right|=0 \tag{10}
\end{equation*}
$$

Let

$$
\begin{equation*}
{ }_{1} F_{0}(a ;-\mathbf{S}, \mathbf{T})=\int_{\mathcal{O}(p)}\left|\mathbf{I}+\mathbf{S} \mathbf{H T} \mathbf{H}^{\top}\right|^{-a}(d \mathbf{H}) \tag{11}
\end{equation*}
$$

the equation is well defined for all $\mathbf{S}$ and $\mathbf{T}$ being positive definite. Without loss of generality, we may take both $\mathbf{S}$ and $\mathbf{T}$ are positive diagonal matrices. Then the joint density function of eigenvalues $f_{1}, \ldots, f_{p}$ is of the form

$$
\begin{equation*}
\phi_{\omega}\left(f_{1}, \cdots, f_{p}\right)=|\boldsymbol{\Omega}|^{-\frac{1}{2} n_{1}}{ }_{1} F_{0}\left(\frac{1}{2}\left(n_{1}+n_{2}\right) ;-\mathbf{\Omega}^{-1}, \mathbf{F}\right) \phi_{*}\left(f_{1}, \ldots, f_{p}\right) \tag{12}
\end{equation*}
$$

where

$$
\begin{equation*}
\phi_{*}\left(f_{1}, \ldots, f_{p}\right)=\frac{\pi^{\frac{1}{2} p^{2}} \Gamma_{p}\left(\frac{1}{2}\left(n_{1}+n_{2}\right)\right)}{\Gamma_{p}\left(\frac{1}{2} n_{1}\right) \Gamma_{p}\left(\frac{1}{2} n_{2}\right) \Gamma_{p}\left(\frac{1}{2} p\right)}|\mathbf{F}|^{\frac{1}{2}\left(n_{1}-p-1\right)} \prod_{i<j}\left(f_{i}-f_{j}\right) \tag{13}
\end{equation*}
$$

Theorem 4. ${ }_{1} F_{0}\left(\frac{1}{2} n ;-\mathbf{\Omega}^{-1}, \mathbf{F}\right)$ is $M T P_{2}$, where $n=n_{1}+n_{2}$.
Proof. Please note that $\left.{ }_{1} F_{0}\left(\frac{1}{2} n ;-\mathbf{\Omega}^{-1}, \mathbf{F}\right)=\int_{\mathcal{O}(p)}\left|\mathbf{I}+\mathbf{\Omega}^{-1} \mathbf{H F} \mathbf{H}^{\top}\right|^{-n / 2}(d \mathbf{H})=|\boldsymbol{\Omega}|^{n / 2} \int_{\mathcal{O}(p)} \right\rvert\, \boldsymbol{\Omega}+$ $\left.\mathbf{H F H}^{\top}\right|^{-n / 2}(d \mathbf{H})$, thus

$$
\frac{\partial \log _{1} F_{0}\left(\frac{1}{2} n ;-\mathbf{\Omega}^{-1}, \mathbf{F}\right)}{\partial \mathbf{F}}=\frac{\int_{\mathcal{O}(p)}-\frac{1}{2} n\left|\mathbf{\Omega}+\mathbf{H F H}^{\top}\right|^{-\frac{1}{2} n}\left(\mathbf{\Omega}+\mathbf{H F H}^{\top}\right)^{-1}(d \mathbf{H})}{\int_{\mathcal{O}(p)}\left|\mathbf{\Omega}+\mathbf{H F H}^{\top}\right|^{-\frac{1}{2} n}(d \mathbf{H})}
$$

Write $B(\boldsymbol{\Omega}, \mathbf{F})=\int_{\mathcal{O}(p)}\left|\boldsymbol{\Omega}+\mathbf{H F H}^{\top}\right|^{-n / 2}(d \mathbf{H})>0$, then after some calculations

$$
\begin{aligned}
\frac{\partial^{2} \log _{1} F_{0}\left(\frac{1}{2} n ;-\mathbf{\Omega}^{-1}, \mathbf{F}\right)}{\partial \boldsymbol{\Omega} \partial \mathbf{F}} & =\left\{\frac{1}{4} n^{2} \int_{\mathcal{O}(p)}\left|\boldsymbol{\Omega}+\mathbf{H F H}^{\top}\right|^{-\frac{1}{2} n}\left(\boldsymbol{\Omega}+\mathbf{H F H}^{\top}\right)^{-2}(d \mathbf{H}) \times B(\boldsymbol{\Omega}, \mathbf{F})\right. \\
& -\frac{1}{4} n^{2}\left[\left.\int_{\mathcal{O}(p)}|\boldsymbol{\Omega}+\mathbf{H F H}|^{\top}\right|^{-\frac{1}{2} n}\left(\boldsymbol{\Omega}+\mathbf{H F H}^{\top}\right)^{-1}(d \mathbf{H})\right]^{2} \\
& \left.+\left.\frac{1}{2} n \int_{\mathcal{O}(p)}|\mathbf{\Omega}+\mathbf{H F H}|^{\top}\right|^{-\frac{1}{2} n}\left(\mathbf{\Omega}+\mathbf{H F H}^{\top}\right)^{-2}(d \mathbf{H}) \times B(\boldsymbol{\Omega}, \mathbf{F})\right\} / B^{2}(\boldsymbol{\Omega} . \mathbf{F})
\end{aligned}
$$

By Schwartz inequality, we then have

$$
\begin{aligned}
\frac{\partial^{2} \log _{1} F_{0}\left(\frac{1}{2} n ;-\boldsymbol{\Omega}^{-1}, \mathbf{F}\right)}{\partial \boldsymbol{\Omega} \partial \mathbf{F}} & \geq \frac{\frac{1}{2} n \int_{\mathcal{O}(p)}\left|\boldsymbol{\Omega}+\mathbf{H F H}^{\top}\right|^{-\frac{1}{2} n}\left(\boldsymbol{\Omega}+\mathbf{H F H}^{\top}\right)^{-2}(d \mathbf{H})}{B(\boldsymbol{\Omega}, \mathbf{F})} \\
& =\frac{1}{2} n \mathcal{E}_{f}\left[\left(\boldsymbol{\Omega}+\mathbf{H F H}{ }^{\top}\right)^{-2} \mid \mathbf{F} ; \boldsymbol{\Omega}\right] \\
& \geq \mathbf{0}
\end{aligned}
$$

with probability one, where $f(\mathbf{H} \mid \mathbf{F} ; \boldsymbol{\Omega})=\left|\boldsymbol{\Omega}+\mathbf{H F} \mathbf{H}^{\top}\right|^{-n / 2} / B(\boldsymbol{\Omega}, \mathbf{F})$ can be viewed as the conditional density function of $\mathbf{H}$ given $\mathbf{F}$. Please note that the matrix $\mathbf{H F H}^{\top}$ is positive definite with probability one, thus the conditional expectation $\left.\mathcal{E}_{f}\left[(\Omega+\mathbf{H F H})^{\top}\right)^{-2} \mid \mathbf{F} ; \Omega\right]$ is positive definite with probability one. Hence, the theorem follows.

Please note that $\phi_{*}\left(f_{1}, \ldots, f_{p}\right)$ is $M T P_{2}$. Thus, by Theorem 4 we have that the density function $\phi_{\omega}\left(f_{1}, \ldots, f_{p}\right)$ is $M T P_{2}$.

### 4.3. Type ${ }_{0} F_{1}$, Bessel: Noncentral Means with Known Covariance

Suppose that the columns of a matrix $p \times n$ matrix $\mathbf{X}$ are independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ and $\mathrm{E}(\mathbf{X})=\mathbf{M}$. Let $w_{i}$ be the eigenvalues of $\left|\mathbf{X} \mathbf{X}^{\top}-w \boldsymbol{\Sigma}\right|=0$ and $\mathbf{W}=$ $\operatorname{diag}\left(w_{1}, \ldots, w_{p}\right)$, then its density function depends only upon $\Omega=\operatorname{diag}\left(\omega_{1}, \ldots, \omega_{p}\right)$, where $\omega_{i}$ are the eigenvalues of $\left|\mathbf{M} \mathbf{M}^{\top}-\omega \boldsymbol{\Sigma}\right|=0$, and is

$$
\begin{equation*}
\psi_{\omega}\left(w_{1}, \ldots, w_{p} ; \boldsymbol{\Omega}\right)=e^{-\frac{1}{2} \operatorname{tr} \boldsymbol{\Omega}}{ }_{0} F_{1}\left(\frac{1}{2} n ; \frac{1}{4} \boldsymbol{\Omega}, \mathbf{W}\right) \psi_{0}\left(w_{1}, \ldots, w_{p}\right) \tag{14}
\end{equation*}
$$

where

$$
\begin{equation*}
\psi_{0}\left(w_{1}, \cdots, w_{p}\right)=\frac{\pi^{\frac{1}{2} p^{2}}}{2^{\frac{1}{2} p n} \Gamma_{p}\left(\frac{1}{2} n\right) \Gamma_{p}\left(\frac{1}{2} p\right)} e^{-\frac{1}{2} \operatorname{trW}}|\mathbf{W}|^{\frac{1}{2}(n-p-1)} \prod_{i<j}\left(w_{i}-w_{j}\right) \tag{15}
\end{equation*}
$$

Furthermore, from Equation (32) of James [19], we have the inverse Laplace transform

$$
\begin{equation*}
{ }_{0} F_{1}\left(\frac{1}{2} n ; \mathbf{\Omega}, \mathbf{W}\right)=\frac{2^{\frac{1}{2} p(p-1)} \Gamma_{p}\left(\frac{1}{2} n\right)}{(2 \pi i)^{\frac{1}{2} p(p+1)}} \int_{\mathcal{R}(\mathbf{T})>0} e^{t r \mathbf{T}}|\mathbf{T}|^{-\frac{1}{2} n}{ }_{0} F_{0}\left(\mathbf{T}^{-1} \Omega, \mathbf{W}\right)(d \mathbf{T}) . \tag{16}
\end{equation*}
$$

Without loss of generality, we may assume that $\mathbf{T}=\operatorname{diag}\left(t_{1}, \ldots, t_{p}\right)$ and also let $\mathbf{T}^{-\mathbf{1}} \boldsymbol{\Omega}=\mathbf{U}^{-1}$, then the Equation (16) becomes

$$
\begin{align*}
& { }_{0} F_{1}(b ; \boldsymbol{\Omega}, \mathbf{W})  \tag{17}\\
& =\frac{2^{\frac{1}{2} p(p-1)} \Gamma_{p}\left(\frac{1}{2} n\right)}{(2 \pi i)^{\frac{1}{2} p(p+1)}} \int_{\mathcal{R}(\mathbf{U})>0} e^{\operatorname{tr}(\mathbf{\Omega} \mathbf{U})}|\mathbf{\Omega} \mathbf{U}|^{-\frac{1}{2} n}{ }_{0} F_{0}\left(\mathbf{U}^{-1}, \mathbf{W}\right)|\mathbf{\Omega}|^{p}(d \mathbf{U}) \\
& =\frac{2^{\frac{1}{2} p(p-1)} \Gamma_{p}\left(\frac{1}{2} n\right)}{(2 \pi i)^{\frac{1}{2} p(p+1)}} \int_{\mathcal{R}(\mathbf{U})>0} g\left(u_{1}, \ldots, u_{p} ; \mathbf{\Omega}\right)_{0} F_{0}\left(\mathbf{U}^{-1}, \mathbf{W}\right)(d \mathbf{U}),
\end{align*}
$$

where

$$
\begin{equation*}
g\left(u_{1}, \ldots, u_{p} ; \boldsymbol{\Omega}\right)=e^{\sum_{i=1}^{p} \omega_{i} u_{i}} \prod_{i=1}^{p} \omega_{i}^{-\frac{1}{2} n+p} \prod_{i=1}^{p} u_{i}^{-\frac{1}{2} n} \tag{18}
\end{equation*}
$$

Theorem 5. ${ }_{0} F_{1}\left(\frac{1}{2} n ; \boldsymbol{\Omega} \mathbf{W}\right)$ is $M T P_{2}$.
Proof. By Equation (18), after some straightforward calculations, we have

$$
\frac{\partial^{2} \log g\left(u_{1}, \ldots, u_{p} ; \boldsymbol{\Omega}\right)}{\partial \omega_{i} \partial u_{i}}=1>0, \forall i=1, \ldots, p .
$$

Thus, $g\left(u_{1}, \ldots, u_{p} ; \boldsymbol{\Omega}\right)$ is $T P_{2}$ pairwise in $\left(u_{i}, \delta_{i}\right)$. Namely, $g\left(u_{1}, \ldots, u_{p} ; \boldsymbol{\Omega}\right)$ is $M T P_{2}$. By the composition lemma (Proposition 1) of Karlin [10], thus the theorem follows.

Please note that $\psi_{0}\left(w_{1}, \ldots, w_{p}\right)$ is $M T P_{2}$, thus by Theorem 5 , we may conclude that the density function $\psi_{\omega}\left(w_{1}, \ldots, w_{p} ; \boldsymbol{\Omega}\right)$ is $M T P_{2}$.

### 4.4. Type ${ }_{1} F_{1}$, Confluent Hypergeometric: Noncentral Eigenvalues

Suppose that the $p \times n_{1}$ matrix variate $\mathbf{X}$ with $n_{1}$ independent columns distributed as $N(\mathbf{M}, \boldsymbol{\Sigma})$ and $p \times n_{2}$ matrix variate $\mathbf{Y}$ with $n_{2}$ independent columns distributed as $N(\mathbf{0}, \boldsymbol{\Sigma})$ for $n_{1} \geq p$. Let $\omega_{i}^{*}$ be the eigenvalues of $\left|\mathbf{M} \mathbf{M}^{\top}-\omega^{*} \boldsymbol{\Sigma}\right|=0$. Write $\mathbf{\Omega}^{*}=\operatorname{diag}\left(\omega_{1}^{*}, \ldots, \omega_{p}^{*}\right)$, then the joint density function of eigenvalues $f_{i}$ of $\left|\mathbf{X} \mathbf{X}^{\top}-f \mathbf{Y} \mathbf{Y}^{\top}\right|=0$, (i) for $n_{1} \geq p$, is of the form

$$
\begin{align*}
& \phi_{\omega^{*}}^{*}\left(f_{1}, \ldots, f_{p} ; \mathbf{\Omega}^{*}\right)  \tag{19}\\
& =e^{-\frac{1}{2} \operatorname{tr} \boldsymbol{\Omega}^{*}}{ }_{1} F_{1}\left(\frac{1}{2}\left(n_{1}+n_{2}\right) ; \frac{1}{2} n_{2} ; \frac{1}{2} \mathbf{\Omega}^{*},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right) \phi^{*}\left(f_{1}, \ldots, f_{p}\right),
\end{align*}
$$

where

$$
\begin{equation*}
\phi^{*}\left(f_{1}, \ldots, f_{p}\right)=\frac{\pi^{\frac{1}{2} p^{2}} \Gamma_{p}\left(\frac{1}{2}\left(n_{1}+n_{2}\right)\right)}{\Gamma_{p}\left(\frac{1}{2} n_{1}\right) \Gamma_{p}\left(\frac{1}{2} n_{2}\right) \Gamma_{p}\left(\frac{1}{2} p\right)} \frac{|\mathbf{F}|^{\frac{1}{2}\left(n_{1}-p-1\right)}}{|\mathbf{I}+\mathbf{F}|^{\frac{1}{2}\left(n_{1}+n_{2}\right)}} \prod_{i<j}\left(f_{i}-f_{j}\right) . \tag{20}
\end{equation*}
$$

Theorem 6. ${ }_{1} F_{1}\left(\frac{1}{2} n ; \frac{1}{2} n_{2} ; \frac{1}{2} \mathbf{\Omega}^{*},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right)$ is $M R R_{2}$, where $n=n_{1}+n_{2}$.
Proof. Please note that

$$
\begin{aligned}
{ }_{1} F_{0}\left(\frac{1}{2} n ; \mathbf{\Omega}^{*-1},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right) & =\int_{\mathcal{O}(p)}\left|\mathbf{I}-\mathbf{\Omega}^{*-1} \mathbf{H}\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1} \mathbf{H}^{\top}\right|^{-\frac{1}{2} n}(d \mathbf{H}) \\
& =\left|\mathbf{\Omega}^{*}\right|^{\frac{1}{2} n} \int_{\mathcal{O}(p)}\left|\mathbf{\Omega}^{*}-\mathbf{H}\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1} \mathbf{H}^{\top}\right|^{-\frac{1}{2} n}(d \mathbf{H})
\end{aligned}
$$

Let $\mathbf{G}=\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}$ and $C\left(\mathbf{\Omega}^{*}, \mathbf{G}\right)=\int_{\mathcal{O}(p)}\left|\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right|^{-n / 2}(d \mathbf{H})>0$, then

$$
\begin{aligned}
& \frac{\partial^{2} \log _{1} F_{0}\left(\frac{1}{2} n ; \mathbf{\Omega}^{*-1}, \mathbf{G}\right)}{\partial \mathbf{\Omega}^{*} \partial \mathbf{G}} \\
& =-\left\{\frac{1}{4} n^{2} \int_{\mathcal{O}(p)}\left|\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right|^{-\frac{1}{2} n}\left(\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right)^{-2}(d \mathbf{H}) \times C\left(\mathbf{\Omega}^{*}, \mathbf{G}\right)\right. \\
& -\frac{1}{4} n^{2}\left[\int_{\mathcal{O}(p)}\left|\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right|^{-\frac{1}{2} n}\left(\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right)^{-1}(d \mathbf{H})\right]^{2} \\
& \left.+\left.\frac{1}{2} n \int_{\mathcal{O}(p)}\left|\mathbf{\Omega}^{*}-\mathbf{H G H}\right|^{\top}\right|^{-\frac{1}{2} n}\left(\mathbf{\Omega}^{*}-\mathbf{H} \mathbf{H}^{\top}\right)^{-2}(d \mathbf{H}) \times C\left(\mathbf{\Omega}^{*}, \mathbf{G}\right)\right\} / C^{2}\left(\mathbf{\Omega}^{*}, \mathbf{G}\right) .
\end{aligned}
$$

By Schwartz inequality, then

$$
\begin{aligned}
& \frac{\partial^{2} \log _{1} F_{0}\left(\frac{1}{2} n ; \mathbf{\Omega}^{*-1}, \mathbf{G}\right)}{\partial \mathbf{\Omega}^{*} \partial \mathbf{G}} \\
& \leq-\frac{1}{2} n \int_{\mathcal{O}(p)}\left|\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right|^{-\frac{1}{2} n}\left(\mathbf{\Omega}^{*}-\mathbf{H G} \mathbf{H}^{\top}\right)^{-2}(d \mathbf{H}) / C\left(\mathbf{\Omega}^{*}, \mathbf{G}\right) \\
& =-\frac{1}{2} n \mathcal{E}_{g}\left[\left(\mathbf{\Omega}^{*}+\mathbf{H G} \mathbf{H}^{\top}\right)^{-2} \mid \mathbf{G} ; \mathbf{\Omega}^{*}\right] \\
& \leq \mathbf{0}
\end{aligned}
$$

with probability one, where $g\left(\mathbf{H} \mid \mathbf{G} ; \mathbf{\Omega}^{*}\right)=\left|\boldsymbol{\Omega}+\mathbf{H G H}^{\top}\right|^{-n / 2} / C\left(\mathbf{\Omega}^{*}, \mathbf{G}\right)$ can be viewed as the conditional density function of $\mathbf{H}$ given $\mathbf{G}$. Please note that the matrix $\mathbf{H G H}{ }^{\top}$ is positive definite with probability one, thus the conditional expectation $\mathcal{E}_{g}\left[\left(\mathbf{\Omega}^{*}+\mathbf{H G} \mathbf{H}^{\top}\right)^{-2} \mid \mathbf{G} ; \mathbf{\Omega}^{*}\right]$ is positive definite with probability one. Then, we have that ${ }_{1} F_{0}\left(\frac{1}{2} n ; \boldsymbol{\Omega}^{*-1}, \mathbf{G}\right)$ is $M R R_{2}$ in $\left(\boldsymbol{\Omega}^{*}, \mathbf{G}\right)$. Also, it is easy to note that $\mathbf{G}$ is the monotone increasing function of $\mathbf{F}$, thus, ${ }_{1} F_{0}\left(\frac{1}{2} n ; \mathbf{\Omega}^{*-1},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right)$ is $M R R_{2}$ in $\left(\mathbf{\Omega}^{*}, \mathbf{F}\right)$.

Next, consider

$$
\begin{align*}
& { }_{1} F_{1}\left(\frac{1}{2} n, \frac{1}{2} n_{2} ; \mathbf{\Omega}^{*},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right)  \tag{21}\\
& =\frac{2^{\frac{1}{2} p(p-1)} \Gamma_{p}\left(\frac{1}{2} n_{2}\right)}{(2 \pi i)^{\frac{1}{2} p(p+1)}} \int_{\mathcal{R}(\mathbf{T})>0} e^{t r \mathbf{T}}|\mathbf{T}|^{-\frac{1}{2} n_{2}}{ }_{1} F_{0}\left(\frac{1}{2} n ; \mathbf{T}^{-1} \mathbf{\Omega}^{*},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right)(d \mathbf{T})
\end{align*}
$$

Let $\mathbf{T}^{-1} \mathbf{\Omega}^{*}=\mathbf{U}^{-1}$, then

$$
\begin{aligned}
& { }_{1} F_{1}\left(\frac{1}{2} n, \frac{1}{2} n_{2} ; \mathbf{\Omega}^{*},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right) \\
& =\frac{2^{\frac{1}{2} p(p-1)} \Gamma_{p}\left(\frac{1}{2} n_{2}\right)}{(2 \pi i)^{\frac{1}{2} p(p+1)}} \int_{\mathcal{R}(\mathbf{U})>0} e^{\operatorname{tr}\left(\mathbf{\Omega}^{*} \mathbf{U}\right)}\left|\mathbf{\Omega}^{*} \mathbf{U}\right|^{-\frac{1}{2} n_{2}}{ }_{1} F_{0}\left(\frac{1}{2} n ; \mathbf{U}^{-1}, \mathbf{W}\right)\left|\mathbf{\Omega}^{*}\right|^{p}(d \mathbf{U}) \\
& =\frac{2^{\frac{1}{2} p(p-1)} \Gamma_{p}\left(\frac{1}{2} n_{2}\right)}{(2 \pi i)^{p(p+1) / 2}} \int_{\mathcal{R}(\mathbf{U})>0} g^{*}\left(u_{1}, \ldots, u_{p} ; \mathbf{\Omega}^{*}\right)_{1} F_{0}\left(\frac{1}{2} n ; \mathbf{U}^{-1}, \mathbf{W}\right)(d \mathbf{U}),
\end{aligned}
$$

where

$$
\begin{equation*}
g^{*}\left(u_{1}, \ldots, u_{p} ; \mathbf{\Omega}^{*}\right)=e^{\Sigma_{i=1}^{p} \omega_{i}^{*} u_{i}} \prod_{i=1}^{p} \omega_{i}^{*-\frac{1}{2} n_{2}+p} \prod_{i=1}^{p} u_{i}^{-\frac{1}{2} n_{2}} . \tag{22}
\end{equation*}
$$

After some straightforward calculations, we have

$$
\frac{\partial^{2} \log g^{*}\left(u_{1}, \ldots, u_{p} ; \mathbf{\Omega}^{*}\right)}{\partial \omega_{i}^{*} \partial u_{i}}=1>0, u_{i}, \forall i=1, \ldots, p
$$

Thus, $g^{*}\left(u_{1}, \ldots, u_{p} ; \mathbf{\Omega}^{*}\right)$ is $T P_{2}$ pairwise in $\left(u_{i}, \omega_{i}^{*}\right)$. Namely, $g^{*}\left(u_{1}, \ldots, u_{p} ; \mathbf{\Omega}^{*}\right)$ is $M T P_{2}$. By the composition lemma (Proposition 1) of Karlin [10], then the theorem follows.

Similar to arguments in Example 5, we may have that $\phi^{*}\left(f_{1}, \ldots, f_{p}\right)$ is $M R R_{2}$. Thus by Theorem 6, we may conclude that the density function $\phi_{\omega^{*}}^{*}\left(f_{1}, \ldots, f_{p} ; \mathbf{\Omega}^{*}\right)$, which is the product of two $M R R_{2}$ functions (Karlin and Rinott [13]), is $M R R_{2}$.
(ii) For $p \geq n_{1}$ the joint density function of eigenvalues $f_{1}, \ldots, f_{p}$ is of the form

$$
\begin{align*}
& \phi_{2 \omega^{*}}^{*}\left(f_{1}, \ldots, f_{p} ; \boldsymbol{\Omega}^{*}\right)  \tag{23}\\
& =e^{-\frac{1}{2} \operatorname{tr} \mathbf{\Omega}^{*}}{ }_{1} F_{1}\left(\frac{1}{2}\left(n_{1}+n_{2}\right) ; \frac{1}{2} p ; \frac{1}{2} \mathbf{\Omega}^{*},\left(\mathbf{I}+\mathbf{F}^{-1}\right)^{-1}\right) \phi_{2}^{*}\left(f_{1}, \ldots, f_{p}\right),
\end{align*}
$$

where

$$
\begin{align*}
& \phi_{2}^{*}\left(f_{1}, \ldots, f_{p}\right)  \tag{24}\\
& =\frac{\pi^{\frac{1}{2} n_{1}^{2}} \Gamma_{n_{1}}\left(\frac{1}{2}\left(n_{1}+n_{2}\right)\right)}{\Gamma_{n_{1}}\left(\frac{1}{2} n_{1}\right) \Gamma_{n_{1}}\left(\frac{1}{2}\left(n_{1}+n_{2}-p\right)\right) \Gamma_{n_{1}}\left(\frac{1}{2} p\right)} \frac{|\mathbf{F}|^{\frac{1}{2}\left(p-n_{1}-1\right)}}{|\mathbf{I}+\mathbf{F}|^{\frac{1}{2}\left(n_{1}+n_{2}\right)}} \prod_{i<j}\left(f_{i}-f_{j}\right) .
\end{align*}
$$

Similar to arguments in Theorem 6, we may conclude that the density function $\phi_{2 \omega^{*}}^{*}\left(f_{1}, \ldots, f_{p} ; \boldsymbol{\Omega}^{*}\right)$ is $M R R_{2}$.

### 4.5. High-Dimensional Wishart Matrices

Suppose that the columns of a matrix $p \times n$ matrix $\mathbf{X}$ are independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ and $\mathrm{E}(\mathbf{X})=\mathbf{0}$. Let $\mathbf{L}=\operatorname{diag}\left(l_{1}, \ldots, l_{p}\right)$, where $l_{i}$ are the $i$ th largest eigenvalue of $\mathbf{X X}{ }^{\top}$. Let $\boldsymbol{\Sigma}=\sigma^{2} \mathbf{I}$ and $c=\lim _{n \rightarrow \infty} p / n, c \in(0,1)$. When dimension $p$ is large, the limiting distribution
of sample spectral eigenvalues has the well-known Marčenko-Pastur distribution $F_{c, \sigma^{2}}$ (M-P law) with index $c$ and scale parameter $\sigma$, which the density function is of the form

$$
f_{c, \sigma^{2}}(x)=\frac{1}{2 \pi x c \sigma^{2}} \sqrt{\left(b \sigma^{2}-x\right)\left(x-a \sigma^{2}\right)}, a \sigma^{2} \leq x \leq b \sigma^{2}
$$

where $a=(1-\sqrt{c})^{2}$ and $b=(1+\sqrt{c})^{2}$ (Marčenko and Paster [21]). After some algebraic manipulations, we can show that

$$
\begin{aligned}
\frac{\partial^{2} \log f_{c, \sigma^{2}}(x)}{\partial x \partial \sigma^{2}} & =\frac{1}{4 \pi c} \frac{(a+b) x^{2}-4 a b x \sigma^{2}+(a+b) a b \sigma^{4}}{\left(b \sigma^{2}-x\right)^{2}\left(x-a \sigma^{2}\right)^{2}} \\
& \geq 0
\end{aligned}
$$

Thus, the Marčenko-Pastur density function is $T P_{2}$ in $\left(x, \sigma^{2}\right)$, and hence Theorem 1 is applicable.

## 5. Remarks

We adopted the notion of convex ordering of two distributions to prove the ordering of Shannon entropies in Theorem 2. Please note that the notions of convex ordering of two distributions, the monotone likelihood ratio and totally positive of order $2\left(M T P_{2}\right)$ are essentially equivalent. For many density functions such as those discussed in Sections 3 and 4, it seems easier to characterize the $M T P_{2}$ property of density functions than that of convex ordering of distribution functions. In practice, we suggest use the notion of $M T P_{2}$ for the comparisons of Shannon entropy measures.

The difference of two Shannon entropies is an intrinsic distribution measure as we emphasized in this paper. As a result, the Shannon entropy measures can be ordered when two underline distributions have the relationship of convex ordering (i.e., the $M T P_{2}$ ) property. The monotonicity of power function, which is discussed basically under the same distributions but with different ordered parameters set up, comes out to be a special case of the results of the comparison of two Shannon entropies which can be even discussed under two totally different distribution functions.

For the problem of hypothesis testing $H_{0}: F(x)=_{c} G(x)$ against $H_{1}: F(x) \leq_{c} G(x)$, test statistic based on the sample version of the difference of two Shannon entropies enjoys the optimal property. The above hypothesis testing problem is equivalent to the hypothesis problem of testing $H_{0}: H_{F}(u)=$ $u, u \in[0,1]$ against $H_{1}: H_{F}(u)$ is convex on $[0,1]$. Let $U_{1: n}<U_{2: n}<\ldots<U_{n: n}$ be the order statistics from $H_{F}$, then the most powerful test is based on the statistics $\sum_{i=1}^{n} \log h_{F}\left(U_{i: n}\right)$, which is the empirical version of the difference of two Shannon entropies. Since $h_{F}(u)$ is monotone increasing in $u$, thus the test has the property of power monotonicity.

Noncentral chi-square distributions play an important role for statistical inference. For the univariate case, a noncentral chi-square or gamma density which are typically Poisson mixture of central chi-square or gamma type densities. Beyond the monotonicity property, we may aware of the following: let $f_{\alpha}(x, \beta)$ denote the univariate gamma density function with the shape parameter $\alpha$ and the scale parameter $\beta$, and its distribution function is denoted by $F_{\alpha}(x, \beta)$. For the fixed scale parameter gamma densities, van Zwet [6] showed the result of convex ordering of gamma distributions, i.e., if $0<\alpha_{1} \leq \alpha_{2}$ then $F_{\alpha_{2}}(x, \beta) \leq_{c} F_{\alpha_{1}}(x, \beta)$. Namely, $f_{\alpha}(x, \beta)$ is $T P_{2}$ in $(x, \alpha)$ when $\beta$ is fixed. Thus, by Theorem 1 if $0<\alpha_{1} \leq \alpha_{2}$ then $\mathrm{I}_{S}\left(f_{\alpha_{1}}\right) \leq \mathrm{I}_{S}\left(f_{\alpha_{2}}\right)$ when $\beta$ is fixed. Similarly, when $\alpha$ is fixed $f_{\alpha}(x, \beta)$ is $T P_{2}$ in $(x, \beta)$. The above results tell us that the larger degrees of freedom is, the larger Shannon entropy measure is when both noncentralities are kept the same. For any two test statistics (or estimators) which are noncentral chi-square distributed with the same noncentrality, but with different degrees of freedom, then the Pitman efficiency (based on the Kullback-Leibler divergence) makes no difference of comparisons. However, our new measure of the difference of two Shannon entropies is nonzero, which means that the new measure makes a difference of comparisons. For the MANOVA models, the density functions of eigenvalues which are Bessel of multivariate gamma type density or confluent hypergeometric of multivariate beta type density such as those discussed in Sections 4.3 and 4.4,
respectively, our results can be directly applied to the tests with monotone acceptance regions to enjoy the power monotonicity for those density functions with the $M T P_{2}$ (MLR) property.

The conceived mixture model arises also for the finite case. A multi-sample model where for some $G(G>1)$, there are $G$ densities $f_{g}(\mathbf{x}), g=1, \ldots, G$ and a sample of size $n_{g}$ is drawn from the density $f_{g}(\mathbf{x})$, so that the pooled sample relates to the mixture density with $w_{g}=n_{g} / \sum_{g=1}^{G} n_{g}$. For the finite mixture case, it is hard to find out the explicit form of Shannon entropy of mixture distribution, even under the multinormal set-up. Suppose that there are $G$ groups, the random vector $\mathbf{X}$ has the mixture density function $f_{G}^{*}(\mathbf{x})=\sum_{g=1}^{G} w_{g} f_{g}(\mathbf{x})$, where $\sum_{g=1}^{G} w_{g}=1$. Then the distribution function is $F_{G}^{*}(\mathbf{x})=P\{\mathbf{X} \leq \mathbf{x}\}=\int_{-\infty}^{\mathbf{x}} \sum_{g=1}^{G} w_{g} f_{g}(\mathbf{y}) d \mathbf{y}=\sum_{g=1}^{G} w_{g} \int_{-\infty}^{\mathbf{x}} f_{g}(\mathbf{y}) d \mathbf{y}=\sum_{g=1}^{G} w_{g} F_{g}(\mathbf{x})$. The $w_{g}$ are nonnegative and they add up to 1 , so the mixture density represents a convex combination of the component densities. We find that the decomposability of Shannon entropy still holds in a different way for the continuous random vectors.

Theorem 7. $I_{S}\left(f_{G}^{*}\right)=\sum_{g=1}^{G} w_{g} I_{S}\left(f_{g}\right)+\sum_{g=1}^{G} w_{g} K\left(f_{g}, f_{G}^{*}\right)$.
Proof. Please note that

$$
\begin{aligned}
\mathrm{I}_{S}\left(f_{G}^{*}\right) & =\int_{-\infty}^{\infty}-\log f_{G}^{*}(\mathbf{x}) d F_{G}^{*}(\mathbf{x}) \\
& =-\int_{-\infty}^{\infty} \log \left(\sum_{g=1}^{G} w_{g} f_{g}(\mathbf{x})\right) d\left[\sum_{i=1}^{G} w_{i} F_{i}(\mathbf{x})\right] \\
& =-\sum_{g=1}^{G} w_{g} \int_{-\infty}^{\infty} f_{g}(\mathbf{x}) \log \left(\sum_{i=1}^{G} w_{i} f_{i}(\mathbf{x})\right) d \mathbf{x} \\
& =-\sum_{g=1}^{G} w_{g} \int_{-\infty}^{\infty} f_{g}(\mathbf{x}) \log f_{g}(\mathbf{x}) d \mathbf{x}+\sum_{g=1}^{G} w_{g} \int_{-\infty}^{\infty} f_{g}(\mathbf{x}) \log \left(\frac{f_{g}(\mathbf{x})}{\sum_{i=1}^{G} w_{i} f_{i}(\mathbf{x})}\right) d \mathbf{x} \\
& =\sum_{g=1}^{G} w_{g} I_{S}\left(f_{g}\right)+\sum_{g=1}^{G} w_{g} K\left(f_{g}, f_{G}^{*}\right)
\end{aligned}
$$

It follows from Theorem 7 that $\mathrm{I}_{S}\left(f_{G}^{*}\right) \geq \sum_{g=1}^{G} w_{g} \mathrm{I}_{S}\left(f_{g}\right)$. We further note that for a mixture model, the Shannon entropy has the decomposability property similar to well-known Fisher's MANOVA model. Please note that the first term on the right hand side represents an average of the individual entropies (i.e., analogous to the within group sum of squares) while the second term is nonnegative and represents the between group distances.

For a high-dimensional Wishart matrix with an unknown scale times the identity matrix being the population covariance matrix, the property of $T P_{2}$ property (i.e., the $M L R$ property) for the limiting empirical density function of eigenvalues is studied in the Section 4.5. Although the Marčenko-Pastur equation (Silverstein [22]) provides the link of limiting empirical spectral distribution, $F$, to the limiting behavior of the population spectral distribution, $H$, one can expect to retrieve the information of $H$ from $F$. However, the difficulty lies within the fact that the relationship between $H$ and $F$ is entangled. Whether the result studied in Section 4.5 can be extended to the general population covariance matrix case or not remains to be clarified. Another focus of random matrices in the literature is related to the Gaussian divisible ensembles. The Gaussian divisible ensembles are matrices of the form $\mathbf{H}_{t}=e^{-t / 2} \mathbf{H}_{0}+\left(1-e^{-t}\right)^{1 / 2} \mathbf{H}_{G}$ where $t>0$ is a parameter, $\mathbf{H}_{0}$ is a Wigner matrix and $\mathbf{H}_{G}$ is an independent Gaussian orthogonal ensemble matrix. The eigenvalue distribution of the Gaussian divisible ensembles are the same as that of the solution of a matrix valued Ornstein-Uhlenbeck process $\mathbf{H}_{t}$ for any time $t \geq 0$. The dynamics of the eigenvalues of $\mathbf{H}_{t}$ is given by the system of so-called Dyson's Brownian motion (Dyson [23]). The treatment of the sample covariance matrix is analogous, but the formulas change slightly (see Erdős, L. et al. [24], for the details). It is interesting to further
study that: under the general model, whether the corresponding limiting spectral density function of eigenvalues, obtained via the convolution of the limiting density function of eigenvalues of initial matrix with the Marčenko-Pastur density function, is $T P_{2}$ or not. We pose those problems as a project of future study.

Author Contributions: Conceptualization, M.-T.T., F.-J.H. and C.-H.T.; methodology, M.-T.T., F.-J.H. and C.-H.T.; validation, M.-T.T., F.-J.H. and C.-H.T.; formal analysis, M.-T.T., F.-J.H. and C.-H.T.; writing-original draft preparation, M.-T.T., F.-J.H. and C.-H.T.; writing-review and editing, M.-T.T.
Funding: The APC was funded by Academia Sinica, Taiwan, R.O.C.
Acknowledgments: The authors woud like to thank the reviewers for their helpful comments.
Conflicts of Interest: The authors declare no conflict of interest.

## References

1. Karlin, S.; Rinott, Y. Entropy inequalities for class of probability distributions I. The univariate case. Adv. Appl. Probab. 1981, 13, 93-112. [CrossRef]
2. Karlin, S.; Rinott, Y. Entropy inequalities for class of probability distributions II. The Multivariate case. Adv. Appl. Probab. 1981, 13, 325-351. [CrossRef]
3. Zografos, K.; Nadarajah, S. Expressions for Renyi and Shannon entropies for multivariate distributions. Stat. Probab. Lett. 2005, 71, 71-84. [CrossRef]
4. Tsai, M.-T. The ordering of Shannon entropies. Stat. Sin. 2017, 27, 1725-1729. [CrossRef]
5. Perlman, M.; Olkin, I. Unbiasedness of invariant tests for MANOVA and other multivariate problems. Ann. Stat. 1980, 8, 1326-1341. [CrossRef]
6. van Zwet, W.R. Convex Transformations of Random Variables; Mathematisch Centrum: Amsterdam, The Netherlands, 1964.
7. Barlow, R.E.; Doksum, K.A. Isotonic tests for convex orderings. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume I; University of California Press: Berkeley, CA, USA, 1972; pp. 293-323.
8. Gastwirth, J.L. A general definition of the Lorenz curve. Econometrica 1971, 39, 1037-1039. [CrossRef]
9. Shaked, M.; Shanthikumar, J. Stochastic Orders; Springer: New York, NY, USA, 2007.
10. Karlin, S. Total Positivity; Stanford University Press: Stanford, CA, USA, 1968.
11. Tong, Y.L. The Multivariate Normal Distribution; Springer: New York, NY, USA, 1990.
12. Karlin, S.; Rinott, Y. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. J. Multivar. Anal. 1980, 10, 467-498. [CrossRef]
13. Karlin, S.; Rinott, Y. Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rule distributions. J. Multivar. Anal. 1980, 10, 499-516. [CrossRef]
14. Gupta, R.D.; Richards, D.S.P. Multivariate Liouville distributions. J. Multivar. Anal. 1987, 23, 233-256. [CrossRef]
15. Cox, D.R. Regression models and life-tables. J. R. Stat. Soc. Ser. B 1972, 34, 187-220. [CrossRef]
16. Dumitriu, I. Eigenvalue Statistics for Beta-Ensembles. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2003.
17. Peddada, S.D.; Richards, D.S.P. Entropy inequality for some multivariate distributions. J. Multivar. Anal. 1991, 39, 202-208. [CrossRef]
18. Dykstra, R.L.; Hewett, J.E. Positive dependence of the roots of a Wishart matrix. Ann. Stat. 1978, 6, 235-238. [CrossRef]
19. James, A.T. The distribution of the latent roots of the covariance matrix. Ann. Math. Stat. 1960, 31, 874-882. [CrossRef]
20. Muirhead, R.J. Aspects of Multivariate Statistical Theory; John Wiley \& Sons: New York, NY, USA, 1982.
21. Marčenko, V.A.; Paster, L.A. Distribution of eigenvalues for some sets of random matrices. Math. USSR-Sbornik 1967, 1, 457-483. [CrossRef]
22. Silverstein, J.W. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. J. Multivar. Anal. 1995, 55, 331-339. [CrossRef]
23. Dyson, F.J. A Brownian-motion model for the eigenvalues of a random matrix. J. Math. Phys. 1962, 3, 1191-1198. [CrossRef]
24. Erdős, L.; Schlein, B.; Yau, H.-T.; Yin, J. The local relaxation flow approach to university of the local statistics for random matrices. Ann. Inst. Henri Poincaré Probab. Stat. 2012, 48, 1-46. [CrossRef]
