Ann. Inst. Statist. Math.
Vol. 46, No. 1, 165–177 (1994)

ON THE JOINT DISTRIBUTION OF STUDENTIZED ORDER STATISTICS

TEA-YUAN HWANG[1] AND CHIN-YUAN HU[2]

[1] *Institute of Statistics, National Tsing Hua University, Hsinchu 30043, Taiwan, R.O.C.*
[2] *Department of Business Education, National Changhua University of Education, Changhua 50058, Taiwan, R.O.C.*

(Received April 9, 1992; revised February 15, 1993)

Abstract. In this paper, the joint distribution of some special linear combinations of the (internally) studentized order statistics are derived for both normal and exponential populations; the exact relationship between their pdf's is also obtained. The exact sampling distributions of studentized extreme deviation statistic, which has been proposed by Pearson and Chandra Sekar (1936, *Biometrika*, **28**, 308–320), are derived for these two populations. An application to the most powerful location and scale invariant test is discussed briefly.

Key words and phrases: Exponential distribution, normal distribution, studentized order statistics.

1. Introduction

The problem of testing for outlying observations is of considerable importance in applied statistics, and the proper treatment of outliers has long been a subject for study (see, Barnett and Lewis (1984) and David (1981) and the references therein). For testing the significance of the smallest (or largest) observation in a sample of size n from a normal population, Pearson and Chandra Sekar (1936) proposed the studentized extreme deviation statistic, and Grubbs (1950) obtained its exact sampling distribution for the normal population by using two well-known transformations. The joint distribution of (internally) studentized order statistics for the normal population is still unknown. In this paper, a very simple and concise expression for exact sampling distribution of some special linear functions of these studentized order statistics will be derived for both normal and exponential populations. It can be used to obtain the distribution of the studentized order statistics.

Let the proposed nonlinear transformations as follows:

$$
\begin{aligned}
t_i = & \left[\frac{n-i+1}{(n-1)(n-i)}\right]^{1/2} \\
& \cdot \left[\frac{y_i - \bar{y}_n}{s_n} + \frac{1}{n-i+1} \sum_{k=1}^{i-1} \frac{y_k - \bar{y}_n}{s_n}\right], \qquad 1 \le i \le n-2, \\
w_1 = & \ \bar{y}_n, \\
w_2 = & \ s_n,
\end{aligned} \tag{1.1}
$$

where $\bar{y}_n = \sum_{i=1}^n y_i/n$, and $s_n^2 = \sum_{i=1}^n (y_i - \bar{y}_n)^2/(n-1)$ and $y_1 \le y_2 \le \cdots \le y_n$. Note that the summation in (1.1) will be taken as zero for $i = 1$.

The key point of this paper is the inverse relationship of the nonlinear transformations given in (1.1). The inverse transformations of (1.1), its derivation will be presented in Theorem 2.1, are as follows:

$$
\begin{aligned}
\frac{y_i - \bar{y}_n}{s_n\sqrt{n-1}} &= \left[\frac{n-i}{n-i+1}\right]^{1/2} \cdot t_i \\
&\quad - \sum_{k=1}^{i-1} \frac{t_k}{[(n-k)(n-k+1)]^{1/2}} \qquad 1 \le i \le n-2, \\
\frac{y_{n-1} - \bar{y}_n}{s_n\sqrt{n-1}} &= -\sum_{k=1}^{n-2} \frac{t_k}{[(n-k)(n-k+1)]^{1/2}} - [f_{n-2}/2]^{1/2}, \\
\frac{y_n - \bar{y}_n}{s_n\sqrt{n-1}} &= -\sum_{k=1}^{n-2} \frac{t_k}{[(n-k)(n-k+1)]^{1/2}} + [f_{n-2}/2]^{1/2}
\end{aligned} \tag{1.2}
$$

where $f_{n-2} = 1 - t_1^2 - \cdots - t_{n-2}^2$.

In Section 2, we discuss some properties of these transformations and then use those results to find their Jacobian which makes the derivation of the pdf of the related $T_1, \ldots, T_{n-2}$ statistics possible. A similar transformation that can be used to deal with exponential population is also given.

In Section 3, the joint pdf's of the studentized order statistics for both normal and exponential populations are obtained; the exact relationship between their pdf's is also established. An application to the most powerful location and scale invariant test is briefly discussed in this section.

In Section 4, the exact sampling distributions of the studentized extreme deviation statistic are derived for both the normal and exponential populations.

Finally, we note that the approach given here can be extended to deal with a larger class of distributions; for example, when the original random vector follows spherical distributions (see Muirhead (1982), p. 37) or is uniformly distributed over a positive simplex in $\mathbb{R}^n$ (see Aitchison (1982, 1985)).

2. Properties of the proposed transformations

In this section, some properties of the set of nonlinear transformations given in (1.1) are discussed, and then we show that the transformation is a one-to-one correspondence between its domain and range except possibly for a set of n-dimensional Lebesgue measure zero. A similar transformation that can be used to deal with the exponential population is also given.

In the following, the variables $y_1, \ldots, y_n$ are assumed not all equal and $y_1 \leq \cdots \leq y_n$ and $n \geq 3$. First, we note that the studentized order variables $(y_1 - \bar{y}_n)/s_n, \ldots, (y_n - \bar{y}_n)/s_n$ are clearly bounded, for example $(n-1)^{1/2}$ is an upper bound for all these studentized order variables. The sharpest bounds for each of the studentized ordered variables were provided in Theorems 3.8 and 3.9 of Arnold and Balakrishnan ((1989), pp. 48–49). As a consequence, we have the following lemmas. Note that their s_n^2 is $(n-1)/n$ times the s_n^2 used in the present discussion.

LEMMA 2.1. *(1) $-1 \leq t_1 \leq -1/(n-1)$, (2) $t_1 = -1$ if and only if $y_1 < y_2 = \cdots = y_n$, (3) $t_1 = -1/(n-1)$ if and only if $y_1 = \cdots = y_{n-1} < y_n$.*

LEMMA 2.2. *(1) $0 \geq t_i \geq -[n/i(n-i+1)]^{1/2}$, $2 \leq i \leq n-2$, (2) the maximum value of t_i can be attained if and only if $y_i = y_{i+1} = \cdots = y_n$, and (3) the minimum of t_i can be attained if and only if $y_1 = \cdots = y_i < y_{i+1} = \cdots = y_n$.*

A non-trivial lower bound for $\bar{y}_n/s_n$ which will be used in Theorem 2.2 below is presented as the following

LEMMA 2.3. *For $0 \leq y_1 \leq \cdots \leq y_n$, we have $\bar{y}_n/s_n \geq 1/\sqrt{n}$ and the equality holds if and only if $0 = y_1 = \cdots = y_{n-1} < y_n$.*

PROOF. The proof of this lemma is straightforward and therefore omitted here. □

LEMMA 2.4.

$$\sum_{i=1}^{k} t_i^2 = \sum_{i=1}^{k} \left[\frac{y_i - \bar{y}_n}{\sqrt{n-1} \cdot s_n}\right]^2 + \frac{1}{n-k}\left[\sum_{i=1}^{k} \frac{y_i - \bar{y}_n}{\sqrt{n-1} \cdot s_n}\right]^2, \qquad 1 \leq k \leq n-2.$$

PROOF. For convenience, let us define

$$a_k = \sum_{i=1}^{n-k} c_i^2 + \frac{1}{k}\left[\sum_{i=1}^{n-k} c_i\right]^2,$$

$$b_k = \left[\frac{k+1}{k}\right]^{1/2} \cdot \left[c_{n-k} + \frac{1}{k+1}\sum_{i=1}^{n-k-1} c_i\right],$$

$1 \leq k \leq n-1$, where $c_1, \ldots, c_{n-1}$ are arbitrary real variables, and set $a_0 = 0$ and $b_{n-1} = \sqrt{n} \cdot c_1/\sqrt{n-1}$. Then we have $a_k = b_k^2 + a_{k+1}$, $a_k = \sum_{i=k}^{n-1} b_i^2$, $1 \leq k \leq n-1$.

Let's choose $c_i = (y_i - \bar{y}_n)/\sqrt{n-1} \cdot s_n$. Then $b_{n-i} = t_i$, $1 \le i \le n-2$, and Lemma 2.4 is established. □

LEMMA 2.5. *Let* $f_i = 1 - t_1^2 - \cdots - t_i^2$, $1 \le i \le n-2$. *Then the following two conditions are equivalent.*

$$\begin{aligned}
&\text{(A)} \quad -1 \le t_1 \le -1/(n-1), \\
&\left[\frac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1} \le t_k, \qquad 2 \le k \le n-2, \\
&t_{n-2} \le -(f_{n-2}/3)^{1/2}, \qquad f_{n-2} \ge 0. \\
&\text{(B)} \quad -1 \le t_1 \le -1/(n-1), \\
&\max\left\{\left[\frac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2}\right\} \le t_k \le -f_{k-1}^{1/2}/(n-k).
\end{aligned}$$

PROOF. $t_{n-2} \le -(f_{n-2}/3)^{1/2}$ and $f_{n-2} = f_{n-3} - t_{n-2}^2$ gives $t_{n-2} \le -f_{n-3}^{1/2}/2$, and combining the last inequality with $\sqrt{2}t_{n-3} \le t_{n-2}$, we have $\sqrt{2}t_{n-3} \le -f_{n-3}^{1/2}/2$ and $t_{n-3} \le -f_{n-4}^{1/2}/3$ since $f_{n-3} = f_{n-4} - t_{n-3}^2$. For $2 \le k \le n-3$, we have

$$\left[\frac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1} \le t_k \le -f_{k-1}^{1/2}/(n-k) \tag{2.1}$$

gives

$$\left[\frac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1} \le -f_{k-1}^{1/2}/(n-k).$$

By the same process, furthermore $t_{k-1} \le -f_{k-2}^{1/2}/(n-k+1)$, since $f_{k-1} = f_{k-2} - t_{k-1}^2$, and finally for $2 \le k \le n-3$

$$t_k \le -f_{k-1}^{1/2}/(n-k).$$

Next, $f_{n-2} \ge 0$ implies $f_{n-3} \ge t_{n-2}^2$ or $t_{n-2} \ge -f_{n-3}^{1/2}$, thus $f_k \ge 0$ gives $t_k \ge -f_{k-1}^{1/2}$ for $2 \le k \le n-2$ by the same process. Combining with (2.1) we have for $2 \le k \le n-2$

$$\max\left\{\left[\frac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2}\right\} \le t_k \le -f_{k-1}^{1/2}/(n-k).$$

Hence condition (A) implies condition (B). The converse is obvious, and Lemma 2.5 is established. □

Using the above lemmas, we prove the following main theorem which makes the proposed transformations useful in the derivation of exact sampling distributions of the related statistics $T_1, \ldots, T_{n-2}$. For notational convenience, the vector $(t_1, \ldots, t_{n-2}, w_1, w_2)$ is denoted by $\boldsymbol{t}$.

THEOREM 2.1. *Let $t_1, \ldots, t_{n-2}$ and w_1, w_2 be defined as in* (1.1) *and let $f_i = 1 - t_1^2 - \cdots - t_i^2$, $1 \le i \le n-2$. Then there exists a one-to-one correspondence between the domain $D_n(\boldsymbol{y})$ and the range $R_n(\boldsymbol{t})$ except for a set of n-dimensional Lebesgue measure zero, where*

$$D_n(\boldsymbol{y}) = \{\boldsymbol{y} : y_1 \le \cdots \le y_n\} \tag{2.2}$$

and

$$R_n(\boldsymbol{t}) = \left\{ \boldsymbol{t} : \begin{array}{c} -1 \le t_1 \le -1/(n-1) \\ \max\left\{ \left[\dfrac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2} \right\} \\ \le t_k \le -f_{k-1}^{1/2}/(n-k) \\ 2 \le k \le n-2,\ w_1 \in \mathbb{R},\ w_2 \ge 0 \end{array} \right\}. \tag{2.3}$$

Furthermore, the inverse transformation is given as in (1.2), *and the absolute value of Jacobian is*

$$|J| = \sqrt{n} \cdot (n-1)^{(n-1)/2} \cdot w_2^{n-2} \cdot f_{n-2}^{-1/2}. \tag{2.4}$$

PROOF. The set of transformations can be divided into the following three sets of transformations:

$$v_i = y_i - \bar{y}_n, \qquad 1 \le i \le n-2, \qquad v_{n-1} = \bar{y}_n, \qquad v_n = s_n$$

and

$$u_i = \left[\frac{n-i+1}{n-i}\right]^{1/2} \left[v_i + \frac{1}{n-i+1} \sum_{k=1}^{i-1} v_k \right], \qquad 1 \le i \le n-2,$$
$$u_{n-1} = v_{n-1}, \qquad u_n = v_n$$

and

$$t_i = \frac{u_i}{\sqrt{(n-1)u_n}}, \qquad 1 \le i \le n-2, \qquad w_1 = u_{n-1}, \qquad w_2 = u_n.$$

Thus their Jacobians can be respectively computed as follows:

$$J_1 = n(n-1)s_n/(y_n - y_{n-1}),$$
$$J_2 = (2/n)^{1/2},$$
$$J_3 = (n-1)^{(n-2)/2} \cdot w_2^{n-2}$$

and the multiplicative theorem of Jacobians and Lemma 2.4 with $k = n-2$ yields the absolute value of Jacobian given in (2.4). Similarly, the inverse transformation (1.2) can be obtained by using the inverse relationship of the above three transformations and Lemma 2.4.

The proposed transformations actually is a one-to-one correspondence between $D_n(\boldsymbol{y})$ and $R_n(\boldsymbol{t})$ except for a set of n-dimensional Lebesgue measure zero since the Jacobian (2.4) is not equal to zero on $D_n(\boldsymbol{y})$ except that $\bar{y}_n = 0$ or $s_n = 0$ or $f_{n-2} = 0$.

To show the onto part, it follows from Lemma 2.4 with $k = n-2$ that

$$\sum_{i=1}^{n-2} t_i^2 = 1 - \frac{(y_{n-1} - y_n)^2}{2(n-1) \cdot s_n^2} \leq 1$$

or equivalently $f_{n-2} \geq 0$. Note that $f_{n-2} = 0$ if and only if $y_{n-1} = y_n$. By inverse transformation we have for $2 \leq i \leq n-2$,

$$(2.5) \quad \begin{aligned} y_i - y_{i-1} &= \sqrt{n-1} \cdot w_2 \cdot \left[\left[\frac{n-i}{n-i+1} \right]^{1/2} \cdot t_i - \left[\frac{n-i+2}{n-i+1} \right]^{1/2} \cdot t_{i-1} \right], \\ y_{n-1} - y_{n-2} &= \sqrt{n-1} \cdot w_2 \cdot \left[-\frac{3 \cdot t_{n-2}}{\sqrt{6}} - (f_{n-2}/2)^{1/2} \right], \\ y_n - y_{n-1} &= \sqrt{n-1} \cdot w_2 \cdot (2 \cdot f_{n-2})^{1/2}. \end{aligned}$$

Since w_2 is non-negative, thus we have $y_1 \leq y_2 \leq \cdots \leq y_n$ if and only if $y_k - y_{k-1} \geq 0$ for $2 \leq k \leq n$; therefore we have from (2.5) and Lemma 2.2

$$\begin{aligned} &\left[\frac{n-i+2}{n-i} \right]^{1/2} \cdot t_{i-1} \leq t_i, \qquad 2 \leq i \leq n-2, \\ &t_{n-2} \leq -(f_{n-2}/3)^{1/2}. \end{aligned}$$

Combining the above arguments and Lemmas 2.1 and 2.5, the desired result follows and the proof of Theorem 2.1 is completed. □

From Lemma 2.3 and the same arguments as that of Theorem 2.1, we have the following theorem.

THEOREM 2.2. *Let $w_3 = \bar{y}_n / s_n$ and w_1 and t_i's, $1 \leq i \leq n-2$ be as defined in (1.1). Then, there exists a one-to-one correspondence between $D_n^*(\boldsymbol{y})$ and $R_n^*(\boldsymbol{t})$ except for a set of n-dimensional Lebesgue measure zero, where*

$$(2.6) \qquad D_n^*(\boldsymbol{y}) = \{\boldsymbol{y} : 0 \leq y_1 \leq \cdots \leq y_n\}$$

and

$$(2.7) \qquad R_n^*(\boldsymbol{t}) = \left\{ \boldsymbol{t} : \begin{array}{c} \max\left\{ -1, \dfrac{-\sqrt{n} \cdot w_3}{n-1} \right\} \leq t_1 \leq \dfrac{-1}{n-1} \\ \max\left\{ \left[\dfrac{n-k+2}{n-k} \right]^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2} \right\} \\ \leq t_k \leq -f_{k-1}^{1/2}/(n-k) \\ 2 \leq k \leq n-2, \ w_1 \geq 0, \ w_3 \geq 1/\sqrt{n} \end{array} \right\}.$$

Furthermore, the inverse transformation is exactly given as in (1.2), and the absolute value of Jacobian is

$$|J| = \sqrt{n} \cdot (n-1)^{(n-1)/2} \cdot w_1^{n-1} \cdot w_3^{-n} \cdot f_{n-2}^{-1/2}.$$

3. The joint pdf of studentized order statistics

In this section, we derive the exact joint pdf of linear combinations of the studentized order statistics defined in Section 1 for both normal and exponential populations; the relationship between their pdf's is also established. The approach given below can be extended to deal with a larger class of distributions, for example, when the original random vector follows spherical distributions or is uniformly distributed over a positive simplex in $\mathbb{R}^n$. Finally, an application to the most powerful location and scale invariant test is briefly discussed.

The part of linear combination of the studentized order statistics $T_1, \ldots, T_{n-2}$ given as in the non-linear transformations (1.1) is equivalent to the studentized order statistics $(Y_{(1)} - \bar{Y}_n)/S_n, \ldots, (Y_{(n)} - \bar{Y}_n)/S_n$. This fact can be seen from the following two conditions:

$$\sum_{i=1}^{n}(Y_{(i)} - \bar{Y}_n)/S_n = 0, \qquad \sum_{i=1}^{n}(Y_{(i)} - \bar{Y}_n)^2/S_n^2 = n-1.$$

First, we will show that the joint distribution of $(T_1, \ldots, T_{n-2})$ is uniformly distributed over a subset of the unit sphere in $\mathbb{R}^{n-1}$ under the normal parent distribution. This remarkable property can be seen from the following theorem.

THEOREM 3.1. *Let $Y_1, \ldots, Y_n$ be iid random variables from standard normal distribution, and let $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ be their order statistics. Then*

$$T_i = \left[\frac{n-i+1}{(n-1)(n-i)}\right]^{1/2}\left[\frac{Y_{(i)} - \bar{Y}_n}{S_n} + \frac{1}{n-i+1}\sum_{k=1}^{i-1}\frac{Y_{(k)} - \bar{Y}_n}{S_n}\right], \tag{3.1}$$

$1 \le i \le n-2$, have joint pdf as follows

$$\frac{n! \cdot \Gamma\left(\dfrac{n-1}{2}\right)}{2 \cdot \pi^{(n-1)/2} \cdot f_{n-2}^{1/2}} \cdot I_{R^*(t_1,\ldots,t_{n-2})} \tag{3.2}$$

where I is the indicator function and

$$R^*(t_1, \ldots, t_{n-2}) = \left\{ (t_1, \ldots, t_{n-2}) : \begin{array}{c} -1 \le t_1 \le -1/(n-1), \\ \max\left\{\left[\dfrac{n-k+2}{n-k}\right]^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2}\right\} \\ \le t_k \le -f_{k-1}^{1/2}/(n-k), \\ 2 \le k \le n-2 \end{array} \right\} \tag{3.3}$$

and $f_i = 1 - t_1^2 - \cdots - t_i^2$, $1 \le i \le n-2$.

PROOF. By Theorem 2.1, we have the joint pdf of $(T_1, \ldots, T_{n-2}, W_1, W_2)$ as following

$$\frac{n! \cdot \sqrt{n} \cdot (n-1)^{(n-1)/2} \cdot w_2^{n-2}}{f_{n-2}^{1/2} \cdot (2\pi)^{n/2}} \exp\left\{\frac{-1}{2}[(n-1)w_2^2 + n \cdot w_1^2]\right\}$$

where $(t_1, \ldots, t_{n-2}) \in R^*(t_1, \ldots, t_{n-2})$, $-\infty < w_1 < +\infty$ and $w_2 > 0$. Since $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n (y_i - \bar{y}_n)^2 + n \cdot \bar{y}_n^2$. Theorem 3.1 follows immediately by integrating out w_1 and w_2. □

A simple and concise expression of the joint distribution of $(T_1, \ldots, T_{n-2})$ for the normal population is given in the following Corollary 3.1; its proof is straightforward and omitted here. For notational convenience, we define a statistic T_{n-1} using (3.1) with $i = n-1$, and it can be shown that $T_1^2 + \cdots + T_{n-1}^2 = 1$.

COROLLARY 3.1. *When the parent distribution is normal, the joint distribution of* $T_1, \ldots, T_{n-1}$ *given in* (3.1) *is uniformly distributed over the subset* A *of the unit sphere in* $\mathbb{R}^{n-1}$, *where*

$$(3.4) \quad A = \left\{ (t_1, \ldots, t_{n-1}) : \begin{array}{l} t_1^2 + \cdots + t_{n-1}^2 = 1, \\ \left(\dfrac{n-k+2}{n-k}\right)^{1/2} \cdot t_{k-1} \le t_k \le 0, \ 2 \le k \le n-1 \end{array} \right\}.$$

It is well-known that the uniform distribution on the unit sphere

$$A_{n-1} = \{(t_1, \ldots, t_{n-1}) : t_1^2 + \cdots + t_{n-1}^2 = 1\}$$

in $\mathbb{R}^{n-1}$ is the unique distribution on A_{n-1} which is invariant under orthogonal transformations (see, for example, Muirhead (1982), p. 37), and that the surface area of the unit sphere A_{n-1} in $\mathbb{R}^{n-1}$, denoted by $|A_{n-1}|$, is given by

$$|A_{n-1}| = \frac{2 \cdot \pi^{(n-1)/2}}{\Gamma\left(\dfrac{n-1}{2}\right)}.$$

Similar result of the surface area of the subset A of the unit sphere in $\mathbb{R}^{n-1}$, given as in (3.4), can be computed from Theorem 3.1; that is

$$|A| = \frac{2 \cdot \pi^{(n-1)/2}}{n!\Gamma\left(\dfrac{n-1}{2}\right)} = \frac{1}{n!}|A_{n-1}|.$$

For convenience, we give the following notation.

DEFINITION. The uniform distribution over the subset A of the unit sphere in $\mathbb{R}^{n-1}$ is denoted by $\sigma_{n-1}(A)$, where the subset A is given as in (3.4).

Similar approach, as mentioned before, can be used to deal with a larger class of distributions by using the main Theorem 2.1. For example, it can be shown that Theorem 3.1 and Corollary 3.1 still hold, when the original random vector $(Y_1, \ldots, Y_n)$ follows a spherical distribution (see, for example, Muirhead (1982), p. 34).

Next, we will deal with non-normal parent distributions by means of Theorem 2.2. In order to avoid complications, we confine ourselves to the case when parent distribution is exponential. Now, by the same approach as mentioned before, we have

THEOREM 3.2. *Let $Y_1, \ldots, Y_n$ be iid random variables from the exponential distribution with parameter θ, and the statistics T_i, $1 \le i \le n-2$, be defined as in (3.1). Then, the joint pdf of $(T_1, \ldots, T_{n-2})$ is given by*

$$\frac{b_n(-t_1)^{-(n-1)}}{(1 - t_1^2 - \cdots - t_{n-2}^2)^{1/2}} \cdot I_{R^*(t_1, \ldots, t_{n-2})} \tag{3.5}$$

where the set $R^*(t_1, \ldots, t_{n-2})$ is given as in (3.3), and

$$b_n = \frac{(n!)^2}{n^{(n+2)/2} \cdot (n-1)^{(n+1)/2}}. \tag{3.6}$$

PROOF. From Theorem 2.2, we have the joint pdf of $(T_1, \ldots, T_{n-2}, W_1, W_3)$ as following

$$\frac{\theta^n \cdot n! \cdot \sqrt{n} \cdot (n-1)^{(n-1)/2} \cdot w_1^{n-1}}{w_3^n \cdot f_{n-2}^{1/2}} \cdot \exp\{-n\theta w_1\} \cdot I_{R_n^*(t)}$$

where $R_n^*(\boldsymbol{t})$ is given as in (2.7). Thus Theorem 3.2 is established by integrating out w_1 and w_3. Note that $w_3 \ge -(n-1)t_1/\sqrt{n}$ by (2.7). □

A concise expression of Theorem 3.2 is the following corollary, its proof is straightforward and therefore is omitted.

COROLLARY 3.2. *Define T_{n-1} by (3.1) with $i = n-1$ and assume that the conditions of Theorem 3.2 hold. Then the joint pdf of $(T_1, \ldots, T_{n-1})$ is*

$$b_n^* \cdot (-t_1)^{-(n-1)} \cdot d\sigma_{n-1}(A) \tag{3.7}$$

where the constant b_n^* is

$$b_n^* = \frac{2 \cdot n! \cdot \pi^{(n-1)/2}}{n^{(n+2)/2} \cdot (n-1)^{(n+1)/2} \cdot \Gamma\left(\dfrac{n-1}{2}\right)}. \tag{3.8}$$

By a similar approach, it can be shown that Theorem 3.2 and Corollary 3.2 still hold, when the original random vector is uniformly distributed over a positive simplex in $\mathbb{R}^n$.

Finally, we establish directly from Corollaries 3.1 and 3.2 an exact relationship between the joint pdf's of $(T_1, \ldots, T_{n-1})$ for the normal and exponential populations.

THEOREM 3.3. *Let $f^{(N)}$ and $f^{(E)}$ be the joint pdf's of $(T_1, \ldots, T_{n-1})$ for the normal and exponential populations, respectively. Then, we have*

$$f^{(E)}(t_1, \ldots, t_{n-1}) = b_n^* \cdot (-t_1)^{-(n-1)} \cdot f^{(N)}(t_1, \ldots, t_{n-1}) \tag{3.9}$$

where the constant b_n^* is given as in (3.8).

Theorem 3.3 can be used to find the most powerful location and scale invariant test of normality against the exponential alternative. This most powerful invariant test is already obtained by Uthoff (1970). His method of derivation is an application of a technique originated by Stein (1956) and further developed by Wijsman (1967), Hájek and Sidak (1967), and Koehn (1970). Here we provide an alternative derivation of the test. The advantage of our approach is that the exact sampling distributions of various test statistics can be found out directly for some cases from Theorem 2.1.

For testing normality against the exponential alternative, it can be shown that $(T_1, \ldots, T_{n-1})$ is a maximal invariant for the group of translation and scale changes, and hence the following corollary follows immediately from the Neyman-Pearson lemma and Theorem 3.3.

COROLLARY 3.3. (Uthoff (1970)) *For testing normality against exponential alternative, the most powerful location and scale invariant test is based on the T_1-statistic.*

The exact sampling distributions of the T_1 statistic under both null and alternative hypotheses will be discussed in the next section; and the complete descriptions of our approach is deferred to a future article.

4. The marginal pdf of T_1

In this section, we discuss the sampling distributions of the T_1 statistic for both normal and exponential populations.

For the normal population case, the exact sampling distribution of T_1 already provided in the works of Pearson and Chandra Sekar (1936) and Grubbs (1950) and Barnett and Lewis (1984). Incidentally, Shapiro and Wilk (1972) and Stephens (1978) gave a test for exponentiality by using statistic W_E, which is equal to the statistic T_1^2 in our notations. The exact sampling distribution of W_E (or equivalently T_1) for exponential population is still unknown (see, Stephens (1978), p. 33). This distribution will be given below (see, Theorem 4.2).

First, we note that the sampling distribution of T_1 for the exponential population case can be derived directly from the normal population case by means of the following relationship,

$$f_{T_1}^{(E)}(t_1) = b_n^* \cdot (-t_1)^{-(n-1)} \cdot f_{T_1}^{(N)}(t_1) \tag{4.1}$$

where $f_{T_1}^{(N)}$ and $f_{T_1}^{(E)}$ are the pdf of T_1 for normal and exponential populations, respectively, and the constant b_n^* is given as in (3.8). The relationship (4.1) follows immediately from Theorem 3.3 by integrating out $t_2, \ldots, t_{n-1}$ in (3.9).

For the sake of completeness, in the following we will derive somewhat new formulations of the exact sampling distributions of T_1 for both normal and exponential populations; the formulation of the normal population case is essentially different from that of Grubbs (1950) and Barnett and Lewis (1984).

For that purpose, we define a function G_n^* as follows: Let $G_3^*(t_1) = (1-t_1^2)^{-1/2}$ for $-1 < t_1 < -1/2$ and zero otherwise; and for $n \geq 4$,

$$G_n^*(t_1) = \int_{R_n^*(t_1)} \cdots \int f_{n-2}^{-1/2} dt_2 \cdots dt_{n-2} \tag{4.2}$$

for $-1 < t_1 < -1/(n-1)$ and zero otherwise, where $R_n^*(t_1)$ is the set by integrating out $t_2, \ldots, t_{n-2}$ over $R^*(t_1, \ldots, t_{n-2})$, given as in (3.3), for each fixed t_1.

The function G_n^* defined above plays an important role in the derivation of the sampling distributions of T_1 for both the normal and exponential populations. It is clear that the function G_n^* is well-defined and continuous on the open interval $(-1, -1/(n-1))$. The following lemma gives an iterative relationship for the function G_n^*.

LEMMA 4.1. *Let $G_n^*(t_1)$ be defined as in (4.2), and let $G_3^*(t_1) = (1-t_1^2)^{-1/2}$ for $-1 < t_1 < -1/2$ and zero otherwise. Then, for $n \geq 3$*

$$G_{n+1}^*(t_1) = (1-t_1^2)^{(n-3)/2} \cdot \int_{g_n(t_1)}^{-1/(n-1)} G_n^*(x)dx, \qquad -1 < t_1 < -1/n,$$

and zero otherwise, where

$$g_n(t_1) = \max\left\{ \left[\frac{n+1}{n-1}\right]^{1/2} \cdot \frac{t_1}{\sqrt{1-t_1^2}}, -1 \right\}. \tag{4.3}$$

PROOF. Use the definition of G_{n+1}^* and apply the transformation $(t_1, t_2, \ldots, t_{n-1}) \to (t_1, x_2, \ldots, x_{n-1})$, where $x_i = t_i/(1-t_1^2)^{1/2}$, $2 \leq i \leq n-1$, which has the Jacobian $|J| = (1-t^2)^{(n-2)/2}$, and the result follows by an application of Fubini's theorem and a convenient change of variables. □

From Theorem 3.1 and the definition of G_n^* the pdf of T_1 for the normal population can be rewritten as

$$f_{T_1}^{(N)}(t_1) = \frac{n! \cdot \Gamma\left(\frac{n-1}{2}\right)}{2 \cdot \pi^{(n-1)/2}} \cdot G_n^*(t_1). \tag{4.4}$$

The following Theorems 4.1 and 4.2 give the formulations of the pdf of T_1 for both the normal and exponential populations, respectively. For notational convenience, the suffix in $f_{T_1}^{(N)}$ and $f_{T_1}^{(E)}$ will be omitted.

THEOREM 4.1. (Normal case) *Let f_n be the pdf of T_1 for the sample size n. When the population is normal, for $n \geq 3$,*

$$f_{n+1}(t_1) = \frac{(n+1)\cdot\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\cdot\Gamma\left(\frac{n-1}{2}\right)}\cdot(1-t_1^2)^{(n-3)/2}\cdot\int_{g_n(t_1)}^{-1/(n-1)} f_n(x)dx,$$

$-1 < t_1 < -1/n$, and zero otherwise, where $g_n(t_1)$ is given as in (4.3). The initial pdf for $n = 3$ is $f_3(t_1) = (3/\pi)(1-t_1^2)^{-1/2}$, for $-1 < t_1 < -1/2$ and zero otherwise.

PROOF. This theorem follows from Lemma 4.1 and (4.4). □

THEOREM 4.2. (Exponential case) *Let f_n be the pdf of T_1 for the sample of size n from the exponential population. For $n \geq 3$,*

$$f_{n+1}(t_1) = \frac{(n-1)^{(n+1)/2}\cdot(1-t_1^2)^{(n-3)/2}}{(n+1)^{(n-1)/2}\cdot(-t_1)^n}\cdot\int_{g_n(t_1)}^{-1/(n-1)} (-x)^{n-1}\cdot f_n(x)dx,$$

$-1 < t_1 < -1/n$, and zero otherwise, where $g_n(t_1)$ is given as in (4.3). The initial pdf for $n = 3$ is $f_3(t_1) = (1/\sqrt{3})t_1^{-2}\cdot(1-t_1^2)^{-1/2}$, for $-1 < t_1 < -1/2$ and zero otherwise.

PROOF. This theorem follows from (4.1) and Theorem 4.1. □

Finally, we note that Theorems 4.1 and 4.2 still hold, when the original random vector follows a spherical distribution and the uniform distribution over a positive simplex in $\mathbb{R}^n$, respectively. The proofs are essentially the same as given here.

Acknowledgements

The authors would like to thank the referees for very careful reading of the original version, which resulted in substantial improvements.

REFERENCES

Aitchison, J. (1982). The statistical analysis of compositional data (with discussion), *J. Roy. Statist. Soc. Ser. B*, **44**, 139–177.

Aitchison, J. (1985). A general class of distributions on the simplex, *J. Roy. Statist. Soc. Ser. B*, **47**, 136–146.

Arnold, B. C. and Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics*, 43–54, Springer, New York.

Barnett, V. and Lewis, T. (1984). *Outliers in Statistical Data*, 2nd ed., Wiley, New York.

David, H. A. (1981). *Order Statistics*, 2nd ed., Wiley, New York.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations, *Ann. Math. Statist.*, **21**, 27–58.

Hajek, J. and Sidak, F. (1967). *Theory of Rank Tests*, 45–52, Academic Press, New York.

Koehn, U. (1970). Global cross sections and the densities of maximal invariants, *Ann. Math. Statist.*, **41**, 2045–2056.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, 32–40, Wiley, New York.

Pearson, E. S. and Chandra Sekar, C. (1936). The efficiency of statistical tools and a Criterion for the rejection of outlying observations, *Biometrika*, **28**, 308–320.

Shapiro, S. S. and Wilk, M. B. (1972). An analysis of variance test for exponential distribution (complete samples), *Technometrics*, **14**, 355–370.

Stein, C. (1956). Some problems in multivariate analysis I, Tech. Report #6, Dept. of Statistics, Stanford University.

Stephens, M. A. (1978). On the W test for exponential with origin known, *Technometrics*, **20**, 33–35.

Uthoff, V. A. (1970). An optimum test property of two well-known statistics, *J. Amer. Statist. Assoc.*, **65**, 1597–1600.

Wijsman, R. A. (1967). Cross-sections of orbits and their application to densities of maximal invariants, *Proc. Fifth Berkeley Symp. on Math. Statist. Prob.*, **1**, 389–400.