#1748

A CHARACTERIZATION OF MULTIVARIATE ℓ_1–NORM SYMMETRIC DISTRIBUTIONS*

Bi–Qi Fang** and Kai–Tai Fang***

Institute of Applied Mathematics
Academia Sinica
P.O. Box 2734, Beijing
China

Abstract. Let $\underset{\sim}{z}$ be an $n \times 1$ interchangeable random vector and $z_{(1)} \leq \dots \leq z_{(n)}$ be its order statistics. Let $u_i = (n-i+1)(z_{(i)}-z_{(i-1)})$, $i=1,\dots,n$ with $z_{(0)} = 0$ and $\underset{\sim}{u} = (u_1,\dots,u_n)'$. The main result is that $\underset{\sim}{z} \overset{d}{=} \underset{\sim}{u}$ iff $\underset{\sim}{z}$ is a multivariate ℓ_1–norm symmetric distribution.

Keywords: multivariate ℓ_1–norm symmetric distribution, normalized spacing, order statistics, survival function.

*This work was supported by the Science Foundation of Academia Sinica.

**B. Q. Fang is a visiting scholar of the University College London.

***K. T. Fang is a visiting professor of the University of North Carolina at Chapel Hill.

1. Introduction

The exponential distribution has long been an object of study for the statisticians, one of the most interesting problems of which is on the characterization of the exponential distribution. Some known results and study methods are summarized in Azlarov and Volodin [1], where it is shown in Theorem 5.2 that if a nonnegative r.v. ξ has an absolutely continuous distribution function F(x) which is strictly increasing on $[0,\infty)$, then it can be determined by the difference of a pair of adjacent order statistics whether ξ is exponential distributed. As the multivariate extension of the exponential distribution, some families of multivariate ℓ_1–norm symmetric distributions have been introduced and studied in [2] and related papers, in which T_n denotes the family of distributions with survival functions being functions in ℓ_1–norm, i.e.

$$T_n = \{L(\underset{\sim}{z}) : P(z_1 > u_1,\ldots,z_n > u_n) = h(\sum_{i=1}^{n} u_i)\},$$

where $L(\underset{\sim}{z})$ denotes the distribution of $\underset{\sim}{z}$. A special member of T_n is the random vector with i.i.d. exponential components. It turns out that there are many similar properties between multivariate ℓ_1–norm symmetric distributions and the exponential distribution. In this paper it is further shown that multivariate ℓ_1–norm symmetric distributions can be characterized in a way similar to the exponential distribution. The main result is the characterization of multivariate ℓ_1–norm symmetric distributions by the adjacent differences of order statistics. Since there is no restriction on components' independence for multivariate ℓ_1–norm symmetric distributions, this result can be regarded as a generalization of Theorem 5.2 in [1]. Moreover, it extends Theorem 2.1 in [3] as explained below.

2. Main Results

Let $\underset{\sim}{z}^* = (z_1^*,\dots,z_n^*) = (z_{(1)},\dots,z_{(n)})$ be the order statistics of $\underset{\sim}{z}$, i.e. $z_1^* \le \dots \le z_n^*$ and the normalized spacings (N.S.) of $\underset{\sim}{z}$ $u_i = (n-i+1)(z_{(i)}-z_{(i-1)})$, $i=1,\dots,n$ $(z_{(0)} = 0)$. In [3] it is shown that if $\underset{\sim}{z}$ has multivariate ℓ_1–norm symmetric distribution, then its N.S. $\underset{\sim}{u}$ has the same distribution as $\underset{\sim}{z}$, i.e. $\underset{\sim}{u} \overset{d}{=} \underset{\sim}{z}$ (this conclusion holds for every $\underset{\sim}{z} \in T_n$ though in the original paper [3] it is stated for a subset of T_n, the family of scale mixtures of the uniform distribution on the surface of the ℓ_1 unit sphere). In the following theorem it will be proved that this is also a sufficient condition for $\underset{\sim}{z} \in T_n$ when interchangeable random variates (multivariate symmetric distribution) are considered. A main argument is stated separately as a lemma. A corollary on the characterization of T_n by conditional distribution is given at the last. Set

$$R_+^n = \{\underset{\sim}{z} \in R^n : z_i \ge 0,\ i=1,\dots,n\}.$$

All (random) vectors in this paper take values in R_+^n.

<u>Theorem 1</u>. Suppose that $\underset{\sim}{z}$ is an interchangeable random vector with p.d.f. f which is continuous function and $\underset{\sim}{u}$ is the N.S. of $\underset{\sim}{z}$. Then $\underset{\sim}{z} \in T_n$ if and only if $\underset{\sim}{z} \overset{d}{=} \underset{\sim}{u}$.

<u>Proof</u>. We first prove sufficiency. Let

$$S_{\underset{\sim}{i}} = \{\underset{\sim}{z} \in R_+^n : z_{i_1} < \dots < z_{i_n}\}$$

where $\underset{\sim}{i} = (i_1,\dots,i_n)$ is a permutation of $(1,\dots,n)$. Define the transformation $\varphi_{\underset{\sim}{i}}$ in $S_{\underset{\sim}{i}}$:

$$\underset{\sim}{z} \to \underset{\sim}{z}^* = (z_{i_1},\dots,z_{i_n}), \tag{1}$$

the Jacobian of which is 1. Hence $\underset{\sim}{z}^*$ has p.d.f.

$$\sum_{\underset{\sim}{i}} f(\dots,\underset{j\,th}{z_k^*},\dots) = \sum_{\underset{\sim}{i}} f(z_1^*,\dots,z_n^*) = n!f(z_1^*,\dots,z_n^*), \quad 0 \le z_1^* < \dots < z_n^*$$

Now make another transformation

$$u_i = (n-i+1)(z_{(i)} - z_{(i-1)}) \qquad (2)$$

the Jacobian of which is n!. Hence $\underset{\sim}{u}$ has p.d.f.

$$f(u_1/n,\ldots,\sum_{j=1}^{i} u_j/(n-j+1),\ldots,u_1/n + \ldots + u_n)$$

The assumption $\underset{\sim}{u} \overset{d}{=} \underset{\sim}{z}$ leads to

$$f(u_1,\ldots,u_n) = f(u_1/n,\ldots,\sum_{j=1}^{i} u_j/(n-j+1),\ldots,u_1/n+\ldots+u_n) \qquad (3)$$

By Lemma 1 below we have

$$f(u_1,\ldots,u_n) = f(0,\ldots,0,\sum_{i=1}^{n} u_i) \qquad (4)$$

so that $\underset{\sim}{z} \in T_n$.

By making transformations (1) and (2) for p.d.f. of $\underset{\sim}{z} \in T_n$, necessity follows (c.f. [3]).

Lemma 1. Suppose that f is a symmetric function on R_+^n which is continuous in the neighborhood of $\{\underset{\sim}{u} : u_1 = \ldots = u_{n-1} = 0\}$ and satisfies (3). Then (4) holds.

Proof. By iterating (3) m times we get

$$f(\underset{\sim}{u}) = f(\ldots,\sum_{k=1}^{i} A(i,k;m)u_k,\ldots) \qquad (5)$$

where

$$A(i,k;m) = (n-k+1)^{-1} \sum_{j_1=k}^{i} (n-j_1+1)^{-1} \sum_{j_2=j_1}^{i} (n-j_2+1)^{-1} \cdots \sum_{j_{m-1}=j_{m-2}}^{i} (n-j_{m-1}+1)^{-1}$$

$$k \leq i\,,$$

$$A(i,k;m) = 0 \qquad k > i, \quad m=1,2,\ldots \qquad (6)$$

is the coefficient of u_k of the i*th* variable after m*th* iteration.

The following equations hold:

$$A(i,k;m) = A(i-1,k;m) + (n-i+1)^{-1} A(i,k;m-1)$$

$$i = 2,\dots,n, \quad k=1,\dots,i, \qquad (7)$$

$$A(n,k;m) = (n-k+1)^{-1} \sum_{r=k}^{n} A(n,r;m-1)$$

$$= (n-k+1)^{-1} A(n,k;m-1) + (n-k)/(n-k+1)A(n,k+1;m)$$

$$k = 1,\dots,n-1 \qquad (8)$$

Hence $A(i,k;m)$ is an increasing sequence of i and $A(n,k;m)$ is an increasing sequence of k and m respectively. $\{A(i,k;m)\}$ is bounded. If the limit of $A(i,k;m)$ when $m \to \infty$ exists, we denote it by $A(i,k)$. We have

$$A(i,i) = 0 \qquad i = 1,\dots,n-1 \qquad (9)$$

$$A(n,n;m) = A(n,n) = 1 \qquad (10)$$

Let $m \to \infty$ in (7). Then

$$\overline{\lim_m}\ A(i,k;m) = \overline{\lim_m}\ A(i-1,k;m) + \overline{\lim_m}\ A(i,k;m)/(n-i+1)$$

$$\overline{\lim_m}\ A(i,k;m) = (n-i+1)/(n-i)\ \overline{\lim_m}\ A(i-1,k;m), \quad i \neq n$$

So by (9) we have

$$\overline{\lim_m}\ A(i,k;m) = 0 \qquad i = 1,\dots,n-1, \quad k=1,\dots,i$$

which leads to

$$A(i,k) = 0 \qquad i=1,\dots,n-1, \quad k=1,\dots,i$$

Since $A(n,k;m)$ is an increasing bounded sequence, $A(n,k)$ exists. Let $m \to \infty$ in (8) we have

$$A(n,1) = A(n,2) = \dots = A(n,n) = 1$$

Finally, let $m \to \infty$ in (5) then (4) holds as a result of the continuity of f.

Corollary 1. Under the condition of Theorem 1, $\underset{\sim}{z} \in T_n$ if and only if N.S. of (z_1,z_2) and (z_1,z_2) have the same distribution given $(z_3,\dots,z_n)$.

Proof. Necessity is obvious. We now prove sufficiency. Let $h(z_3,\dots,z_n)$ be p.d.f. of $(z_3,\dots,z_n)$. By Theorem 1 the conditional distribution of (z_1,z_2) given $(z_3,\dots,z_n)$ can be denoted by $g(z_1+z_2 | z_3,\dots,z_n)$. Hence

$$f(z_1,\dots,z_n) = g(z_1+z_2;\ z_3,\dots,z_n)\, h(z_3,\dots,z_n) = f(0,\, z_1+z_2,\, z_3,\dots,z_n)$$

which together with the symmetry of f leads to (4) by iteration. Thus $\underset{\sim}{z} \in T_n$.

If components' independence is imposed on $\underset{\sim}{z} \in T_n$, then $z_1,\dots,z_n$ are exponential distributed [2]. This can be seen from the fact that

$$h(x+y) = P(z_1 > x,\, z_2 > y) = P(z_1 > x)\, P(z_2 > y) = h(x)h(y)$$

$$h(0) = 1$$

imply

$$h(t) = e^{-at},\ a > 0$$

Therefore, the above theorem and corollary describe the characterization of exponential distribution in the case of $z_1,\dots,z_n$ being i.i.d. absolutely continuous and nonnegative random variables.

References

[1] Azlarov, T.A. and Volodin, N.A. (1986) *Characterization problems associated with the exponential distribution* , Springer–Verlag, New York.

[2] Fang, K.T. and Fang, B.Q. (1987) Some families of multivariate symmetric distributions related to exponential distribution, *J. Mult. Anal.*, *23*.

[3] Fang, B.Q. and Fang, K.T. (1988) Order statistics of multivariate symmetric distributions related to exponential distribution and their applications, to appear in *Chinese Journal of Applied Probability and Statistics*, *4*.