JOURNAL OF MULTIVARIATE ANALYSIS 33, 1–16 (1990)

Elliptical Symmetry and Exchangeability with Characterizations

ABDULHAMID A. ALZAID

King Saud University, Saudi Arabia

C. RADHAKRISHNA RAO*

University of Pittsburgh

AND

D. N. SHANBHAG

University of Sheffield, Sheffield S37RH, United Kingdom

Communicated by the Editors

We establish certain general characterization results on elliptically symmetric distributions and exchangeable random variables. These results yield, in particular, the results given earlier by Maxwell, Bartlett, Kingman, Ali, Smith, Arnold and Lynch, and several others.

1. INTRODUCTION

Several important properties of elliptically symmetric distributions (e.s.d.s) or, in particular, of spherically symmetric distributions (s.s.d.s) are in some sense extended versions of certain properties of multivariate normal distributions. Many of these properties appear to be of statistical relevance with applications especially in Bayesian inference. A review of the literature on e.s.d.s is contained in a recent paper by Chmielewski [6].

There are several ways of defining e.s.d.s or s.s.d.s. However, we shall restrict ourselves in this discussion to the following definitions.

Received August 2, 1988.

AMS 1980 subject classifications: primary 62H10; secondary 60H05, 62E10.

Key words and phrases: elliptically and spherically symmetric distributions, exchangeable random variables, de Finetti's theorem, Skitovič–Darmois theorem, Schoenberg representation for spherical distributions.

* Now at the Pennsylvania State University, University Park, PA 16802.

0047-259X/90 $3.00

A distribution on R^n is said to be an e.s.d. if its characteristic function $\phi(\mathbf{t})$ is a real function of $\mathbf{t}\Sigma\mathbf{t}'$ for some non-negative definite real symmetric matrix Σ. Furthermore, a distribution on R^n is said to be an s.s.d. if its characteristic function $\phi(\mathbf{t})$ is a real function of $\|\mathbf{t}\|$ ($\|\cdot\|$ stands for the usual norm).

From the work of Schoenberg [18] and subsequent authors it is known that an n-vector random variable $\mathbf{X}$ (assumed clearly to be a row) has an e.s.d. as defined above if and only if it has the representation

$$\mathbf{X} \stackrel{d}{=} T\mathbf{U}C, \tag{1.1}$$

where C is an $n \times n$ real matrix such that $C'C = \Sigma$, $\mathbf{U}$ is an n-vector random variable that is uniformly distributed on $\{\mathbf{x} \in R^n: \|\mathbf{x}\| = 1\}$ and T is a non-negative real random variable independent of $\mathbf{U}$. It is implicit here that the distribution of $\mathbf{X}$ in question depends on C only through Σ. An s.s.d. is obviously a special case of e.s.d. Indeed, any n-vector random variable $\mathbf{X}$ has an s.s.d. if and only if it satisfies (1.1) with $C = I$. In the present paper, we use the notation $\mathscr{E}_n(\Sigma)$ for the class of e.s.d.s defined by (1.1). The class of spherical distributions is denoted by $\mathscr{E}_n(I)$.

Maxwell [15], Bartlett [4], Kac [10], Hartman and Wintner [7], Kelker [12], Nash and Klamkin [16], Ali [1], Arnold and Lynch [3], and several others have characterized the normal distributions in the class $\mathscr{E}_n(I)$, for $n \geq 2$, as the one having properties such as:

1. The components $X_1, X_2, ..., X_n$ of the random variable $\mathbf{X}$ are independently distributed.

2. $\bar{X} = (1/n)(X_1 + \cdots + X_n)$ and $S^2 = (1/(n-1))[(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]$ are independently distributed.

Kingman [13] and, more recently, Smith [21] have characterized certain mixtures of univariate normal distributions via sphericity and exchangeability.

The purpose of the present paper is to give some general results characterizing subclasses of distributions in $\mathscr{E}_n(\Sigma)$. In Section 2, we give results that enable us to provide a unified treatment to the problems considered in the literature cited above, and also to translate the well-known characterizations of normality, such as Darmois–Skitovič theorem and other results, under the assumption of independence of the components of $\mathbf{X}$, based on properties of linear or quadratic statistics, to characterizations of normality in the class $\mathscr{E}_n(I)$. In Section 3, we characterize certain probability distributions through regression properties of statistics based on exchangeable random variables yielding, among other things, certain characterizations of members of $\mathscr{E}_n(I)$ and related distributions. As a consequence of our general results, we obtain the results of Kingman [13] and

Smith [21] and also give various other characterizations of distributions such as mixed gama, Poisson, binomial, and negative binomial distributions.

2. Characterizations of Normality Based on Ellipticity

In this section, we give two general results on the characterization of the normal distribution in the class $\mathscr{E}_n(I)$. These results immediately establish that properties such as

$$E(S^2 \mid \bar{X}) = c \text{ a.s.}, \qquad E(S^2 \mid |\bar{X}|) = c \text{ a.s.}, \qquad E(S \mid \bar{X}) = c \text{ a.s.}$$

$$E(S \mid |\bar{X}|) = c \text{ a.s.}, \qquad E(|\bar{X}| \mid S^2) = c \text{ a.s.}, \qquad E(\bar{X}^2 \mid S^2) = c \text{ a.s.},$$

etc., where each of the c's, appearing above and in other equations of the paper, represents a constant which could take different values in different cases, are characteristic properties of the normal distribution in the class $\mathscr{E}_n(I)$. These results also give several versions of the Darmois–Skitovič theorem under ellipticity or sphericity of the distribution of $\mathbf{X} = (X_1, ..., X_n)$.

Define now the following polynomial functions,

$$Q_1(\mathbf{x}) = \sum\nolimits_1 a_{i_1, ..., i_r} \prod_{r=1}^{n} x_r^{i_r}, \qquad \mathbf{x} = (x_1, ..., x_n) \in R^n \tag{2.1}$$

$$Q_2(\mathbf{x}) = \sum\nolimits_2 b_{i_1, ..., i_r} \prod_{r=1}^{n} x_r^{i_r}, \qquad \mathbf{x} = (x_1, ..., x_n) \in R^n, \tag{2.2}$$

where $\sum_1$ and $\sum_2$ denote summations over $\{(i_1, ..., i_n): i_1 + \cdots + i_n = k\}$ and $\{(i_1, ..., i_n): i_1 + \cdots + i_n = m\}$, respectively, with $k, m \geqslant 1$ and fixed. We shall assume Q_1 and Q_2 to be such that

$$P\{Q_1(\mathbf{U}C) = 0\} \neq 1 \qquad \text{and} \qquad P\{Q_2(\mathbf{U}C) = 0\} \neq 1, \tag{2.3}$$

where $\mathbf{U}$ is an n-vector random variable with uniform distribution on the unit sphere, for some (and hence all) $n \times n$ real C satisfying the condition $C'C = \Sigma$. Note that the assumption (2.3) is equivalent to

$$P\{Q_1(\mathbf{U}C) = 0\} = 0 \qquad \text{and} \qquad P\{Q_2(\mathbf{U}C) = 0\} = 0 \tag{2.4}$$

for some (and hence all) C satisfying the condition $C'C = \Sigma$ (see note A_2 in the Appendix). Then we have the following basic theorem.

THEOREM 1. *Let $b>0$, Q_1 and Q_2 be as defined in* (2.1), (2.2), *and F be a specified nondegenerate member of $\mathscr{E}_n(\Sigma)$ for which the property*

$$E\{|Q_1(\mathbf{X})|^b \mid |Q_2(\mathbf{X})|\} = c \text{ a.s.} \tag{2.5}$$

holds, where $\mathbf{X} \sim F$. Then the property (2.5) *characterizes F except for a change of scale of the random variable involved, in the class of nondegenerate distributions in $\mathscr{E}_n(\Sigma)$. (That is, if there is a nondegenerate r.v. $\mathbf{Y}$ whose distribution belongs to $\mathscr{E}_n(\Sigma)$ and satisfies* (2.5), *then there exists a constant $\lambda>0$ such that $\lambda\mathbf{Y}$ has the distribution F.)*

Proof. Let $\mathbf{X} \sim F \in \mathscr{E}_n(\Sigma)$ and $\mathbf{Y} \sim G \in \mathscr{E}_n(\Sigma)$ both have the property (2.5). By the representation (1.1), $\mathbf{X} =^d T_1 \mathbf{U} C$ and $\mathbf{Y} =^d T\mathbf{U}C$, where T_1 and T are nonnegative r.v.'s. We show that $T =^d \lambda T_1$, which establishes the theorem.

The property (2.5) for $\mathbf{Y}$, with $c=1$, without loss of generality, implies

$$\begin{aligned} &E\{|Q_1(\mathbf{U}C)|^b \, |Q_2(\mathbf{U}C)|^\theta \, T^{bk+m\theta}\} \\ &\quad = E\{|Q_2(\mathbf{U}C)|^\theta \, T^{m\theta}\} \qquad \text{for} \quad \theta \geqslant 0 \end{aligned} \tag{2.6}$$

with $T>0$ a.s. (Note that $P\{T=0\}>0$ is impossible, since it contradicts (2.5) in view of the fact that $P\{Q_i(\mathbf{U}C)=0\}=0$, $i=1,2$. Also, in (2.6) we take $\theta \geqslant 0$ only because it is sufficient for our purpose; indeed Eq. (2.6) is valid for all real θ with possibly infinite value for both sides for some θ values.) In view of the independence of T and $\mathbf{U}$, (2.6) implies

$$\begin{aligned} &E\{|Q_1(\mathbf{U}C)|^b \, |Q_2(\mathbf{U}C)|^\theta\} \, E\{T^{bk+m\theta}\} \\ &\quad = E\{|Q_2(\mathbf{U}C)|^\theta\} \, E\{T^{m\theta}\} \qquad \text{for} \quad \theta \geqslant 0. \end{aligned} \tag{2.7}$$

On taking $\theta = 0, m^{-1}bk, 2m^{-1}bk, \ldots$, successively, (2.7) yields inductively

$$E\{T^{abk}\} < \infty \qquad \text{for all integers } a \geqslant 0$$

(since $E(T^0)=1$, and the identity (2.7) gives $E\{T^{(a+1)bk}\} < \infty$ whenever $E\{T^{abk}\} < \infty$) and hence that $E\{T^\theta\} < \infty$ for all nonnegative real θ. Consequently, we have from (2.7) the identity, with both sides well defined and finite,

$$E\{T^{bk+m\theta}\} = \frac{E\{|Q_2|^\theta\}}{E\{|Q_1|^b \, |Q_2|^\theta\}} \, E(T^{m\theta}), \qquad \theta \geqslant 0, \tag{2.8}$$

where $Q_i = Q_i(\mathbf{U}C)$, $i = 1, 2$. It follows from Lemma A_1 given in the Appendix and the fact that $|Q_2(\mathbf{U}C)|$ is bounded almost surely, that

$$\frac{d}{d\theta}\log\frac{E\{|Q_2|^\theta\}}{E\{|Q_1|^b\,|Q_2|^\theta\}} = \frac{E\{\log|Q_2|\,|Q_2|^\theta\}}{E\{|Q_2|^\theta\}} - \frac{E\{\log|Q_2|\,|Q_1|^b\,|Q_2|^\theta\}}{E\{|Q_1|^b\,|Q_2|^\theta\}} \tag{2.9}$$

$$\to \log\alpha - \log\alpha \qquad \text{as} \quad \theta \to \infty \tag{2.10}$$

with α as the right extremity of the distribution of $|Q_2|$. To establish (2.10), we take ψ of Lemma A_1 to be such that $\psi(z) = \log z$, $z \in (0, \infty)$ and Z as respectively $|Q_2|$ and a r.v. Z with d.f.,

$$E\{|Q_1|^b\, I_{\{|Q_2| \leqslant z\}}\}/E\{|Q_1|^b\}.$$

Then it follows that each of the two quantities in (2.9) tends to $\log\alpha$. From (2.10), we conclude that

$$\frac{d}{d\theta}\log E\{T^{bk+\theta}\} - \frac{d}{d\theta}\log E\{T^\theta\} \to 0 \qquad \text{as} \quad \theta \to \infty. \tag{2.11}$$

Since $E\{T^\theta\}$, $\theta \in [0, \infty)$ is the restriction of a moment generating function to $[0, \infty)$, it is obvious that it is log-convex and hence $d\log E\{T^\theta\}/d\theta$ is increasing on $(0, \infty)$. Consequently, from (2.11), we conclude that

$$\frac{d}{d\theta}\log E\{T^{\theta+\theta'}\} - \frac{d}{d\theta}\log E\{T^\theta\} \to 0 \qquad \text{as} \quad \theta \to \infty \tag{2.12}$$

for every $\theta' \geqslant 0$. If T and T_1 are two positive r.v.'s for which (2.8) is valid, then it follows that $E\{T^\theta\}/E\{(T_1)^\theta\}$ is periodic with period bk for $\theta \geqslant 0$. Since (2.12) is valid for T as well as T_1,

$$\frac{d}{d\theta}\log\frac{E\{T^\theta\}}{E\{(T_1)^\theta\}}$$

is independent of θ for $\theta \in (0, \infty)$ or, equivalently,

$$E\{T^\theta\} = \lambda^\theta E\{(T_1)^\theta\}, \qquad \theta \in [0, \infty) \tag{2.13}$$

for some positive λ. In view of the uniqueness theorem for Mellin transforms, Eq. (2.13) implies $T =^d \lambda T_1$, which proves the theorem. (Incidentally if both T and T_1 satisfy (2.8) we get $\lambda = 1$; however, we have arrived at this situation only because our argument does not take into account scale changes.)

THEOREM 2. *Suppose that* $|Q_1(\mathbf{X})|$ *and* $|Q_2(\mathbf{X})|$ *are independently dis-*

tributed, where $\mathbf{X} \sim N_n(\mathbf{0}, \Sigma) \in \mathscr{E}_n(\Sigma)$. *If the independence property holds for any other nondegenerate* $\mathbf{Y} \sim F \in \mathscr{E}_n(\Sigma)$, *then* F *is the same as* $N_n(\mathbf{0}, \Sigma)$, *except for a change of scale of the random variable.*

Proof. Assume $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ to be independent if $\mathbf{X} \sim N_n(\mathbf{0}, \Sigma)$. It is now sufficient if we show that this condition on $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ with $\mathbf{X}$ having nondegenerate distribution in $\mathscr{E}_n(\Sigma)$ implies that $\mathbf{X}$ has a distribution of the type $N_n(\mathbf{0}, \alpha\Sigma)$ for some α. In view of Eq. (1.1) and the condition that $P(Q_i(\mathbf{U}C) = 0) = 0$, $i = 1, 2$, it follows that the independence of $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ together with the nondegeneracy of $\mathbf{X}$ implies T of (1.1) to be >0 a.s. We then obtain, appealing to the independence of $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ and condition (1.1), that, for s_1, s_2 in some neighbourhood of the origin,

$$\frac{E\{T^{i(s_1+s_2)}\}}{E\{T^{is_1}\}\, E\{T^{is_2}\}} = \frac{E\{|Q_1(\mathbf{U}C)|^{is_1}\}\, E\{|Q_2(\mathbf{U}C)|^{is_2}\}}{E\{|Q_1(\mathbf{U}C)|^{is_1}\, |Q_2(\mathbf{U}C)|^{is_2}\}} \tag{2.14}$$

with both sides well defined. Since the independence of $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ is assumed to be valid for $\mathbf{X}$ with distribution $N_n(\mathbf{0}, \Sigma)$, we can claim that (2.14) is valid for the T corresponding to $N_n(\mathbf{0}, \Sigma)$. Denoting the T for $N_n(\mathbf{0}, \Sigma)$ by T_0, we then see that, for s_1, s_2 in some neighbourhood of the origin,

$$\Psi(s_1 + s_2) = \Psi(s_1)\, \Psi(s_2), \tag{2.15}$$

where

$$\Psi(s) = \frac{E\{T^{is}\}}{E\{T_0^{is}\}}.$$

(Ψ is indeed well defined for all s; however, we require here only the information that it is well defined in a neighbourhood of the origin.) From Lemma 1.5.1 in Kagan, Linnik, and Rao [11] and the fact that $E\{T^{is}\}$ and $E\{T_0^{is}\}$ are characteristic functions, we can then conclude that, for s lying in some neighbourhood of the origin,

$$E\{T^{is}\} = \lambda^{is} E\{T_0^{is}\}$$

with $\lambda > 0$. Since the distribution of $\log(\lambda T_0)$ (i.e., the logarithm of the square root of a certain gamma random variable) is determined uniquely by its moments, we have the distribution of $\log T$ to be determined by its moments and

$$\log T \overset{d}{=} \log(\lambda T_0)$$

or, equivalently,

$$T \stackrel{d}{=} \lambda T_0.$$

Hence we have the required result.

Remark 1. Theorem 1 remains valid if $|Q_2(\mathbf{X})|$ in (2.5) is replaced by $Q_2(\mathbf{X})$, since the identity with $Q_2(\mathbf{X})$ in place of $|Q_2(\mathbf{X})|$ yields the original identity (2.5). Also Theorem 2 remains valid if either $|Q_1(\mathbf{X})|$ is replaced by $Q_1(\mathbf{X})$ or $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ are replaced respectively by $Q_1(\mathbf{X})$ and $Q_2(\mathbf{X})$; this follows because the independence of $|Q_1(\mathbf{X})|$ and $Q_2(\mathbf{X})$ implies the independence of $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$ and also the independence of $Q_1(\mathbf{X})$ and $Q_2(\mathbf{X})$ implies that of $|Q_1(\mathbf{X})|$ and $|Q_2(\mathbf{X})|$.

Remark 2. A result analogous to that of Theorem 2 or its modified version mentioned in Remark 1 remains valid for the uniform distributions on $\{\mathbf{x} \in R^n\colon \|\mathbf{x}\| = \lambda\}$ (with $\lambda > 0$). (Indeed the proof of Theorem 2 illustrates that the result is valid if we replace $N_n(\mathbf{0}, \Sigma)$ by a member of $\mathscr{E}_n(\Sigma)$ that has the corresponding $\log T$ well defined with its distribution determined uniquely by moments; we have not come across any examples of members of $\mathscr{E}_n(\Sigma)$ other than normal and spherically uniform, for which some $Q_1(\mathbf{X})$ and $Q_2(\mathbf{X})$ are independent.) It may also be noted that there exist distributions F which are neither normal nor spherically uniform, for which a property of the type (2.5) is valid. In particular, if $\mathbf{X}$ denotes the subvector containing the first n $(\geqslant 2)$ components of a random vector which has uniform distribution on $\{\mathbf{x} \in R^{n+m}\colon \|\mathbf{x}\| = 1\}$, where $m \geqslant 1$, then

$$E\{(n+m-1)\,X_1^2 + X_2^2 \mid X_2^2\} = 1 \qquad \text{a.s.};$$

however, $\mathbf{X}$ here is neither normal nor spherically uniform. Clearly the $\mathbf{X}$ satisfies (1.1) with $C = I$ and T as the square root of a beta random variable with parameters $n/2$ and $m/2$, respectively. Taking a clue from this example, it is possible to produce more general examples of the type given below supporting our claim:

Example. Let $\mathbf{X} = (X_1, ..., X_n)$ be an n-vector random variable such that

$$\mathbf{X} \stackrel{d}{=} T\mathbf{U}$$

with $\mathbf{U}$ as an n-vector random variable that is uniformly distributed on $\{\mathbf{x} \in R^n\colon \|\mathbf{x}\| = 1\}$ and T as the square root of a beta random variable with parameters $n/2$ and r, respectively, and distributed independently of $\mathbf{U}$, where r is any fixed positive real number. Observe that

$$E\left\{(k+2r)\left(\sum_{i=1}^{k} X_i^2/k\right)+\sum_{i=k+1}^{n} X_i^2 \,\middle|\, \sum_{i=k+1}^{n} X_i^2\right\}=1 \qquad \text{a.s.},$$

where $1 \leqslant k < n$. In the present case, obviously $\mathbf{X}$ is neither spherically uniform nor normal; also, if r is not an integral multiple of $\frac{1}{2}$, we cannot have $\mathbf{X}$ to be of the type considered in the previous example.

Remark 3. It is not true that every characterization of the normal distribution based on n independent identically distributed r.v.'s remains valid in the class of s.s.d.s (on R^n). In particular, the characterization of the normal distribution given by Heyde [9] which is based on the symmetry of a conditional distribution does not, in general, remain valid among s.s.d.s. (A corrected version of Heyde's result appears in Kagan, Linnik, and Rao [11, p. 418].) On the other hand, from Theorems 1 and 2 it follows that there exist characterization properties of normality such as $E(X_1^2+\cdots+X_m^2 \mid X_{m+1}^2+\cdots+X_n^2)=c$ a.s. or independence of $X_1^2+\cdots+X_m^2$ and $X_{m+1}^2+\cdots+X_n^2$, where $1 \leqslant m < n$, in the class of s.s.d.s, which do not happen to be characterization properties of normality in the class of all probability distributions on R^n under the hypothesis that the components of $\mathbf{X}$ are independent and identically distributed.

3. Characterizations Based on Exchangeable Random Variables

Kingman [13] and more recently Smith [21] have characterized certain mixtures of normal distributions based on infinite sequences of exchangeable random variables via the celebrated deFinetti theorem. Smith [21] has also given a statistical motivation to this result, especially in Bayesian inference. In what follows we shall give a theorem implying that characterizations of probability distributions based on sequences of independent and identically distributed random variables of a fairly general type can easily be translated into characterizations of mixtures of probability distributions based on sequences of exchangeable random variables. This result is clearly of potential importance in Bayesian inference and yields several results including those of Kingman and Smith.

Let m be a positive integer and B be a Borel subset of R^m. Further, let $\{F(\cdot \mid \mathbf{b}) : \mathbf{b} \in B\}$ be a family of probability distributions such that $F(x \mid \cdot)$ is Borel measurable on B for each $x \in R$ and $\{(\psi_r^{(1)}, \phi_r^{(1)}, \psi_r^{(2)}, \phi_r^{(2)}) : r \in \Gamma\}$ be a countable family of vectors of real-valued Borel measurable functions on

R^∞ such that for every probability measure on the Borel σ-field of R^∞ yielding all the projection maps to be independent and identically distributed and every $r \in \Gamma$, $(\psi_r^{(1)}, \phi_r^{(1)})$ and $(\psi_r^{(2)}, \phi_r^{(2)})$ are independent and identically distributed. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)$ be a vector of extended real-valued tail Borel measurable functions on R^∞. We have now the following theorem:

THEOREM 3. *Suppose any infinite sequence $\{X_n: n = 1, 2, \ldots\}$ of independent and identically distributed random variables satisfies the condition that*

$$E\{\phi_r^{(1)}(\mathbf{X}) \mid \psi_r^{(1)}(\mathbf{X})\} = 0, \qquad r \in \Gamma \text{ a.s.}, \tag{3.1}$$

where $\mathbf{X} = (X_1, X_2, \ldots)$, only if there exists a $\mathbf{b} \in B$ such that $P\{\boldsymbol{\xi}(\mathbf{X}) = \mathbf{b}\} = 1$ and the df of X_1 is $F(\cdot \mid \mathbf{b})$. Then, if $\{Y_n: n = 1, 2, \ldots\}$ is a sequence of exchangeable random variables, the equation

$$E\{\phi_r^{(1)}(\mathbf{Y})\, \phi_r^{(2)}(\mathbf{Y}) \mid \psi_r^{(1)}(\mathbf{Y}) - \psi_r^{(2)}(\mathbf{Y})\} = 0, \qquad r \in \Gamma \text{ a.s.}, \tag{3.2}$$

where $\mathbf{Y} = (Y_1, Y_2, \ldots)$, is valid only if $\boldsymbol{\xi}(\mathbf{Y}) \in B$ a.s. and conditional upon $\boldsymbol{\xi}(\mathbf{Y})$, the random variables Y_i's are independent and identically distributed with df $F(\cdot \mid \boldsymbol{\xi}(\mathbf{Y}))$ a.s.

Proof (In this proof the identities are to be read as the ones with left-hand side well defined and equal to right-hand side). Equation (3.2) is equivalent to

$$E\{e^{it(\psi_r^{(1)}(\mathbf{Y}) - \psi_r^{(2)}(\mathbf{Y}))}\phi_r^{(1)}(\mathbf{Y})\, \phi_r^{(2)}(\mathbf{Y})\} = 0, \qquad r \in \Gamma,\ -\infty < t < \infty \tag{3.3}$$

(cf. Laha and Lukacs [14, p. 103]; the result in Laha and Lukacs [14] could be viewed as a consequence of the uniqueness theorem corresponding to Fourier transforms of finite signed measures). In view of the deFinetti theorem (see, for example, Olshen [17] for the relevant details), it follows that (3.3), in turn, implies that

$$|E\{e^{it\psi_r^{(1)}(\mathbf{Y})}\phi_r^{(1)}(\mathbf{Y}) \mid \mathscr{T}\}|^2 = 0, \qquad r \in \Gamma \text{ a.s.},\ -\infty < t < \infty, \tag{3.4}$$

where $\mathscr{T}$ is the tail σ-field relative to $\{Y_n: n = 1, 2, \ldots\}$, and hence that

$$E^*\{e^{it\psi_r^{(1)}(\mathbf{Y})}\phi_r^{(1)}(\mathbf{Y}) \mid \mathscr{T}\} = 0, \qquad -\infty < t < \infty,\ r \in \Gamma \text{ a.s.}, \tag{3.5}$$

where E^* is the expectation of $e^{it\psi_r^{(1)}(\mathbf{Y})}\phi_r^{(1)}(\mathbf{Y})$ arrived at via the conditional distribution of $(\psi_r^{(1)}(\mathbf{Y}), \phi_r^{(1)}(\mathbf{Y}))$ given $\mathscr{T}$. Clearly (3.5) is valid in view of the cited result in Laha and Lukacs [14] if and only if

$$E\{\phi_r^{(1)}(\mathbf{Y}) \mid \sigma(\mathscr{T} U \sigma(\psi_r^{(1)}(\mathbf{Y})))\} = 0, \qquad r \in \Gamma, \text{ a.s.} \tag{3.6}$$

In view of the fact that given $\mathscr{T}$, $\{Y_n: n = 1, 2, \ldots\}$ is a sequence of independent identically distributed random variables almost surely, the assumption appearing in the statement of the theorem implies that (3.6) is valid only if $P\{\xi(\mathbf{Y}) \in B \mid \mathscr{T}\} = 1$ a.s. and the conditional distribution of Y_1 given $\mathscr{T}$ is $F(\cdot \mid \xi(\mathbf{Y}))$ a.s. The assertion of the theorem now follows easily.

Remark 4. If $\psi_r^{(1)}, \psi_r^{(2)}$ for each $r \in \Gamma$ in the above theorem are both nonnegative or both nonpositive then the theorem is also valid with $\psi_r^{(1)}(\mathbf{Y}) - \psi_r^{(2)}(\mathbf{Y})$ replaced by $\psi_r^{(1)}(\mathbf{Y}) + \psi_r^{(2)}(\mathbf{Y})$ for some or all $r \in \Gamma$.

Remark 5. If $\psi_r^{(1)}, \psi_r^{(2)}$ are independent of r and Γ is finite, then the theorem remains valid if (3.2) is replaced by

$$\sum_{r \in \Gamma} E\{\phi_r^{(1)}(\mathbf{Y})\, \phi_r^{(2)}(\mathbf{Y}) \mid \psi^{(1)}(\mathbf{Y}) - \psi^{(2)}(\mathbf{Y})\} = 0, \qquad \text{a.s.,} \tag{3.7}$$

where $\psi^{(1)}$ and $\psi^{(2)}$ denote the functions $\psi_r^{(1)}$ and $\psi_r^{(2)}$ that are independent of r. Additionally, if we have $\psi^{(1)}$ and $\psi^{(2)}$ both nonnegative or both nonpositive, then the result in question remains valid when $\psi^{(1)}(\mathbf{Y}) - \psi^{(2)}(\mathbf{Y})$ in (3.7) is replaced by $\psi^{(1)}(\mathbf{Y}) + \psi^{(2)}(\mathbf{Y})$.

Remark 6. The multivariate versions of the above results are clearly valid in view of deFinetti's theorem for vectors.

Remark 7. If the sequence $\{Y_n : n = 1, 2, ...\}$ of exchangeable random variables is taken such that for two distinct real values of θ,

$$E\{e^{\theta(\psi_r^{(1)}(\mathbf{Y}) + \psi_r^{(2)}(\mathbf{Y}))} \mid \phi_r^{(1)}(\mathbf{Y})\, \phi_r^{(2)}(\mathbf{Y}) \mid\} < \infty, \qquad r \in \Gamma,$$

then the above theorem remains valid with the left side of (3.1) replaced by the means of the conditional distributions of $\phi_r^{(1)}(\mathbf{X})$ given $\psi_r^{(1)}(\mathbf{X})$ for $r \in \Gamma$ and simultaneously the left-hand side of (3.2) replaced by the means of the conditional distribution of $\phi_r^{(1)}(\mathbf{Y})\, \phi_r^{(2)}(\mathbf{Y})$ given $\psi_r^{(1)}(\mathbf{Y}) + \psi_r^{(2)}(\mathbf{Y})$ for $r \in \Gamma$.

Remark 8. If it is assumed that $E\{|\phi_r^{(1)}(\mathbf{Y})\, \phi_r^{(2)}(\mathbf{Y})|\} < \infty$ for $r \in \Gamma$, then the results of Theorem 3 and also Remark 7 remain valid with "only if" in two places replaced by "if and only if."

Kingman [13] and more recently Smith [21] have characterized variance mixtures of symmetric normal distributions and mean-variance mixtures of normal distributions on the basis of exchangeability and sphericity. To be more precise, Kingman [13] showed that if $\{Y_n : n = 1, 2, ...\}$ is a sequence of exchangeable random variables such that $(Y_1, ..., Y_4)$ has a spherical distribution, then there exists a nonnegative random variable V such that conditional on V, the Y_n's are independent $N(0, V)$ random variables (a.s.). Smith [21] extended Kingman's ideas to show essentially that if $\{Y_n : n = 1, 2, ...\}$ is a sequence of exchangeable random variables such that $(Y_1, ..., Y_8)$ has a central spherically symmetric distribution, then there exist random variables M, V with $V \geqslant 0$ such that conditional on (M, V), the Y_n's are independent $N(M, V)$ random variables. (It may be noted that in Kingman's and Smith's results degenerate distributions are allowed to be called normal; also it is easily seen that the resulting sequence $\{Y_n\}$ in Kingman's case has for each n the distribution of $(Y_1, ..., Y_n)$ to be spherical and, in Smith's case, for each n the distribution of the vector to be centered spherically symmetric.)

From Box and Hunter [5], it is clear in view of (1.1) that if $(Y_1, ..., Y_4)$ is spherically symmetric, then (for some version)

$$\mu^{(t)}_{1,1,0,0} + \mu^{(t)}_{4,4,0,0} - 6\mu^{(t)}_{4,2,2,0} + 9\mu^{(t)}_{2,2,2,2} = 0, \qquad t \geqslant 0, \tag{3.8}$$

where $\mu^{(t)}_{l_1,l_2,l_3,l_4}$ are the product moments corresponding to the conditional distribution of (Y_1, Y_2, Y_3, Y_4) given $Y_1^2 + \cdots + Y_4^2 = t$. From the definition of the centered spherical symmetry in Smith [21], it follows that if $(Y_1, ..., Y_8)$ is centered spherically symmetric, (3.8) is valid with $\mu^{(t)}_{l_1,l_2,l_3,l_4}$ as the product moments corresponding to the conditional distribution of

$$(Y_1 - Y_2, Y_3 - Y_4, Y_5 - Y_6, Y_7 - Y_8),$$

given

$$(Y_1 - Y_2)^2 + (Y_3 - Y_4)^2 + (Y_5 - Y_6)^2 + (Y_7 - Y_8)^2 = t.$$

We shall now give the following extended versions of the results of Kingman [13] and Smith [21].

COROLLARY 1. *Let $\{Y_n: n = 1, 2, ...\}$ be a sequence of exchangeable random variables. Then, defining $\mu^{(t)}_{l_1,l_2,l_3,l_4}$ to be the product moments corresponding to the conditional distribution of (Y_1, Y_2, Y_3, Y_4) given $Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 = t$, (3.8) is valid if and only if there exists a nonnegative random variable V such that conditional on V, the random variables Y_n are independently distributed $N(0, V)$ random variables (a.s.).*

Proof. The "if" part is trivial. We shall now establish the "only if" part. Take $m = 1$, $\xi_1(\mathbf{x}) = \lim_{n \to \infty} (1/n) \sum_1^n x_i^2$ if the limit exists and equals zero otherwise. Define $\{F(\cdot \mid b)\}$ to be the family of normal probability distribution functions with zero mean. Define

$$\Gamma = \{1, 2\},$$

$$\phi_1^{(1)}(\mathbf{x}) = x_1^4 - 3(x_1 x_2)^2, \qquad \phi_2^{(1)}(\mathbf{x}) = x_1,$$

$$\phi_1^{(2)}(\mathbf{x}) = x_3^4 - 3(x_3 x_4)^2, \qquad \phi_2^{(2)}(\mathbf{x}) = x_3,$$

$$\psi_r^{(1)}(\mathbf{x}) = x_1^2 + x_2^2, \qquad r = 1, 2$$

$$\psi_r^{(2)}(\mathbf{x}) = x_3^2 + x_4^2, \qquad r = 1, 2.$$

$$\begin{aligned}\operatorname{Var}[g_n(x)] &\leqslant \frac{C^2}{n_x} + 4C \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} |W_{ni}(x)\, W_{nj}(x)|\, \psi_{[(i-j)/2]} \\ &\leqslant \frac{C^2}{n_x} + 4C \sup_{i\leqslant n} |W_{ni}(x)| \sum_{j=1}^{n-1} |W_{nj}(x)| \sum_{i=j+1}^{n} \psi_{[(i-j)/2]} \\ &\leqslant \frac{C^2}{n_x} + 4CB \sup_{i\leqslant n} |W_{ni}(x)| \left(\sum_{i=1}^{[n/2]} \psi_i\right) \\ &\to 0 \end{aligned} \tag{9}$$

as $n \to \infty$, since (A2)(b) and (A3) imply $n_x \to \infty$ as $n \to \infty$. ∎

Remark 2.4. (a) It is obvious from inequality (9) that Theorem 2.5 still holds true if we replace assumption $\sum_{m=1}^{\infty} \psi_m < \infty$ by (10), i.e.,

$$\sup_{i\leqslant n} |W_{ni}(x)| \left[\sum_{i=1}^{n} \psi_i\right] \to 0 \quad \text{as} \quad n \to \infty. \tag{10}$$

(b) When $\varepsilon_i^{(n)}$'s are independent r.v.'s, condition (10) holds automatically, since $\psi_i = 0$ for $i \geqslant 1$. Hence, Theorem 2.5 is a natural generalization of Theorem 3 in Georgiev (1988) to the model with dependent observations.

EXAMPLE 2.2. Consider the *k-NN* estimator given in Mack [14] with weights

$$\bar{W}_{ni}(x) = K\left(\frac{x - x_i}{R_n}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{x - x_i}{R_n}\right), \qquad i = 1, 2, \ldots, n \tag{11}$$

where $K(\cdot)$ is a bounded, nonnegative weight function satisfying $K(u) = 0$, for $\|u\| \geqslant 1$. R_n is the Euclidean distance between x and its kth nearest neighbor, and $k = k_n$ satisfies $k_n \to \infty$, $k_n/n \to 0$ as $n \to \infty$. The fixed design points $x_1, \ldots, x_n$ are the same as those in Example 2.1.

It can be shown that $R_n \to 0$, and $nR_n^p \to \infty$, as $n \to \infty$. Therefore, R_n in *k-NN* estimator plays the same role as h_n in Nadaraya–Watson estimator discussed in Georgiev [13]. Assumptions (A2) and (A3) can be verified for $\bar{W}_{ni}(x)$ by using Lemma 2 of Georgiev along the same line as that of the proof of Theorem 1 in Georgiev [13].

We can conclude from the results obtained so far in this section that the *k-NN* estimator is weak and mean square error consistent upon the satisfaction of the other conditions in Corollary 2.4 and Theorem 2.5, respectively. Further, the array of nearest neighbor weights $(\bar{W}_{ni}x))$ is fixed design universally consistent, provided the remaining conditions in Theorem 2.3 are satisfied.

modified version of (3.1) is valid, then X_1 is normal with mean $\xi_1(\mathbf{X})$ and variance $\xi_2(\mathbf{X})$ a.s. The required result is then immediate.

Remark 9. In view of what appears in Laha and Lukacs [14], Kagan, Linnik, and Rao [11], Shanbhag [19, 20], Heller [8], and in several other places, it is obvious that one can arrive at various other characterizations of mixtures of probability distributions on the basis of exchangeable random variables. In particular, the resulting sequence of Corollary 2 is characterized under a mild restriction that $Y_1 Y_2$ is square integrable by the condition that $\{Y_n\}_1^\infty$ is an exchangeable sequence of random variables satisfying for some $k \geqslant 2$,

$$E\{(S_1^2 - \xi_2(\mathbf{Y}))(S_2^2 - \xi_2(\mathbf{Y})) \mid \bar{Y}_1 - \bar{Y}_2\} = 0 \qquad \text{a.s.,}$$

where ξ_2 is as defined in the proof of Corollary 2,

$$\bar{Y}_1 = \frac{1}{k} \sum_{n=1}^{k} Y_n, \qquad \bar{Y}_2 = \frac{1}{k} \sum_{n=1}^{k} Y_{n+k},$$

$$S_1^2 = \frac{1}{k-1} \sum_{n=1}^{k} (Y_n - \bar{Y}_1)^2, \qquad S_2^2 = \frac{1}{k-1} \sum_{n=1}^{k} (Y_{n+k} - \bar{Y}_2)^2.$$

In the same notation and under the same restriction, we have $\{Y_n : n = 1, 2, ...\}$ to be a sequence of exchangeable random variables satisfying

$$E\{(S_1^2 - \bar{Y}_1)(S_2^2 - \bar{Y}_2) \mid \bar{Y}_1 - \bar{Y}_2\} = 0 \qquad \text{a.s.}$$

(or, alternatively, satisfying

$$E\{(S_1^2 - \bar{Y}_1)(S_2^2 - \bar{Y}_2) \mid \bar{Y}_1 + \bar{Y}_2\} = 0 \qquad \text{a.s.} \tag{3.9}$$

when the restriction of square integrability of $Y_1 Y_2$ is replaced by the restriction that Y_n's are nonnegative and the conditional expectations are taken as in Remark 7) if and only if there exists a nonnegative random variable W such that conditional upon W, Y_n's are independent Poisson random variables with mean W a.s. (We allow here the degenerate distribution at zero to be called Poisson; it is also worth pointing out that (3.9) is implied in particular by the condition that Y_n's are such that the conditional distribution of $(Y_1, ..., Y_{2k})$ given $\sum_1^{2k} Y_i$ is multinomial a.s.) Applying the result of Laha and Lukacs [14], it is possible to arrive at analogous characterizations relative to other distributions such as binomial, negative binomial, gamma, and Meixner hypergeometric. (For the definition of a Meixner hypergeometric distribution see Lai [14a].) These characterizations have obviously statistical interpretations of the type given in Smith [21].

Remark 10. In Bayesian inference, the marginal distribution corresponding to a random sample of size n could be viewed as the distribution of any n-members of an infinite sequence of exchangeable random variables. Consequently one could appeal to the results of the present section to characterize well-known marginal distributions here via certain regression properties.

Appendix

A_1. Lemma. *If Z is a positive random variable and ψ is a continuous increasing or decreasing function on $(0, \infty)$ such that $\psi(Z)\, Z^\theta$ and Z^θ are integrable for each $\theta > \theta_0$ $(\geqslant 0)$, then*

$$E\{\psi(Z)\, Z^\theta\}/E\{Z^\theta\} \to \lim_{z \to \alpha^*} \psi(z) \qquad \text{as} \quad \theta \to \infty,$$

where α^* is the right extremity of the distribution of Z. (One could obviously extend the statement of the lemma further; however, the existing form is sufficient for our purposes.)

Proof. There is no loss of generality in assuming ψ to be increasing. Then using the notation $E(Y; A)$ to denote $\int_A Y\, dP$ corresponding to a random variable Y when A is an event, we have, for $z \in (0, \alpha^*)$,

$$\begin{aligned} \lim_{z \to \alpha^*} \psi(z) &\geqslant E\{\psi(Z)\, Z^\theta\}/E\{Z^\theta\} \\ &= \frac{E\{\Psi(Z)(Z/z)^\theta; Z < z\} + E\{\psi(Z)(Z/z)^\theta; Z \geqslant z\}}{E\{(Z/z)^\theta; Z < z\} + E\{(Z/z)^\theta; Z \geqslant z\}} \\ &\geqslant \frac{\psi_1(z, \theta) + \psi(z)}{\psi_2(z, \theta) + 1}, \end{aligned} \qquad (A_1.1)$$

where

$$\psi_1(z, \theta) = E\{\psi(Z)(Z/z)^\theta; Z < z\}/E\{(Z/z)^\theta; Z \geqslant z\}$$

and

$$\psi_2(z, \theta) = E\{(Z/z)^\theta; Z < z\}/E\{(Z/z)^\theta; Z \geqslant z\}.$$

By the Lebesgue dominated convergence theorem and monotone convergence theorem, it follows that

$$\psi_i(z, \theta) \to 0 \qquad \text{as} \quad \theta \to \infty \text{ for } i = 1, 2.$$

Consequently, it follows that the extreme right-hand side of $(A_1.1)$ tends to

$\psi(z)$ as $\theta \to \infty$. Since z is arbitrary, we can then conclude that $(A_1.1)$ and monotonic continuous nature of ψ imply the required result.

A_2. There is no loss of generality if we take $C = I$. If $P(Q_i(\mathbf{U}) = 0) > 0$, then we have either

$$Q_i(u_1, ..., u_{n-1}, \sqrt{1 - u_1^2 - \cdots - u_{n-1}^2}) \equiv 0$$

or

$$Q_i(u_1, ..., u_{n-1}, -\sqrt{1 - u_1^2 - \cdots - u_{n-1}^2}) \equiv 0$$

for $(u_1, ..., u_{n-1})$ lying in some nonempty open subset of $\{\mathbf{x} \in R^{n-1}: \|\mathbf{x}\| \leqslant 1\}$ and hence for all $(u_1, ..., u_{n-1})$ such that $u_1^2 + \cdots u_{n-1}^2 \leqslant 1$. (Note that

$$Q_i(u_1, ..., u_{n-1}, (-1)^j \sqrt{1 - u_1^2 - \cdots - u_{n-1}^2})$$

has a power series expansion for both $j = 1$ and 2 in terms of $u_1, ..., u_{n-1}$ on $\{(u_1, ..., u_{n-1}); u_1^2 + \cdots + u_{n-1}^2 \leqslant 1\}$.) This, in turn, implies that if $P(Q_i(\mathbf{U}) = 0) > 0$, we should have both

$$Q_i(u_1, ..., u_{n-1}, \sqrt{1 - u_1^2 - \cdots - u_{n-1}^2}) \equiv 0$$

and

$$Q_i(u_1, ..., u_{n-1}, -\sqrt{1 - u_1^2 - \cdots - u_{n-1}^2}) \equiv 0$$

for all $(u_1, ..., u_{n-1})$ such that $u_1^2 + \cdots + u_{n-1}^2 \leqslant 1$ (since $Q_i(\mathbf{x}) = 0$ if and only if $Q_i(-\mathbf{x}) = 0$). Consequently, we have that $P(Q_i(\mathbf{U}) = 0) > 0 \Rightarrow P(Q_i(\mathbf{U}) = 0) = 1$.

References

[1] Ali, M. M. (1980). Characterization of the normal distribution among continuous symmetric spherical class. *J. Roy. Statist. Soc. Ser. B* **42** 162–164.

[2] Alzaid, A. A. (1983). *Some Contributions to Characterization Theory.* Ph.D. thesis, University of Sheffield.

[3] Arnold, S. F., and Lynch, J. (1982). On Ali's characterization of spherical normal distribution. *J. Roy. Statist. Soc. Ser. B* **44** 49–51.

[4] Bartlett, M. S. (1934). The vector representation of a sample. *Proc. Cambridge Philos. Soc.* **30** 327–340.

[5] Box, G. E. P., and Hunter, J. S. (1957). Multivariate experimental designs for exploring responses surfaces. *Ann. Math. Statist.* **28** 195–241.

[6] Chmielewski, M. A. (1981). Elliptically symmetric distribution: A review and Bibliography. *Internat. Statist. Rev.* **49** 67–74.

[7] Hartman, P., and Wintner, A. (1940). On the spherical approach to the normal distribution law. *Amer. J. Math.* **62** 759–779.

[8] HELLER, B. (1983). Special functions and characterizations of probability distributions by zero regression properties. *J. Multivariate Anal.* **13** 473–487.

[9] HEYDE, C. C. (1970). Characterization of the normal law by a symmetry of a certain conditional distribution. *Sankhyā Ser. A* **32** 115–118.

[10] KAC, M. (1939). On a characterization of the normal distribution. *Amer. J. Math.* **61** 726–728.

[11] KAGAN, A. M., LINNIK, YU. V., AND RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York.

[12] KELKER, D. (1970). Distribution theory of spherical distributions and location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–430.

[13] KINGMAN, J. F. C. (1972). On random sequences with spherical symmetry. *Biometrika* **59** 492–494.

[14] LAHA, R. G., AND LUKACS, E. (1960). On a problem connected with quadratic regression. *Biometrika* **47** 335–343.

[14a] LAI, C. D. (1982). A characterization of gamma Meixner hypergeometric and negative binomial distributions based on canonical measures. *Ann. Inst. Statist. Math.* **34**, Part A, 359–363.

[15] MAXWELL, J. C. (1860). Illustration of the dynamical theory of gases – Part I. On the motions and collisions of perfectly elastic bodies. *Taylor's Philos. Mag.* **19** 19–32.

[16] NASH, D., AND KLAMKIN, M. S. (1976). A spherical characterization of the normal distribution. *J. Math. Anal. Appl.* **55** 156–158.

[17] OLSHEN, R. (1974). A note on exchangeable sequences. *Z. Warsch. Gebiete* **28** 317–321.

[18] SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *Ann. Math.* **39** 811–841.

[19] SHANBHAG, D. N. (1978). On some problems in distribution theory. *Sankhyā Ser. A* **40** 208–213.

[20] SHANBHAG, D. N. (1979). Some refinements in distribution theory. *Sankhyā Ser. A* **41** 252–262.

[21] SMITH, A. F. M. (1981). On random sequences with centered spherical symmetry. *J. Roy. Statist. Soc. B* **43** 208–209.