RT-MAE-9322

A PREDICTIVISTIC INTERPRETATION TO THE MULTIVARIATE t DISTRIBUTION

by

R.B. Arellano-Valle, H. Bolfarine

and

P.L. Iglesias-Zuazola

Palavras Chaves: Conjugate prior distributions, exponential family, de Finetti's Theorem, orthogonally invariant distributions, multivariate t distribution, inverted Wishart distribution and inverted gamma distribution.
(key words)

Classificação AMS: 62A05
(AMS Classification) 62A15

- Novembro de 1993 -

A PREDICTIVISTIC INTERPRETATION TO THE MULTIVARIATE t DISTRIBUTION

R.B. Arellano-Valle, H. Bolfarine and P.L. Iglesias-Zuazola

Departamento de Estatística - IME
Unviversidade de São Paulo - Caixa Postal 20570
CEP 01492-990 - São Paulo, BRASIL

Abstract

de Finetti type theorems characterize models in terms of invariance. The idea is to begin with observables, postulates of simmetry and then represent the model as mixtures of standard parametric models. If additional conditions are given, then the mixing measure can be determined. Invariance under the action of special groups of orthogonal transformations may give results on the mixtures of parametric normal distributions (Diaconis, Eaton and Lauritzen, 1992). The additional conditions to determining the mixing measure in this case can be obtained following some results in Diaconis and Ylvisaker (1979,1985). Using these results, we obtain a predictivistic characterization of the multivariate t distribution. Also, we give conditions under which the n-dimensional law of sequences of random variables is a location mixture of multivariate t distributions. The results are extended to the case of sequences of random vectors orthogonally invariant.

Key Words and Phrases: Conjugate prior distributions, exponential family, de Finetti's Theorem, orthogonally invariant distributions, multivarite t distribution, inverted Wishart distribution and inverted gamma distribuiton

1. Introduction

A n-dimensional random vector $\mathbf{X}$ is said to have a multivariate t distribution with location parameter $\mathbf{m} = (m_1, \ldots, m_n)' \in \mathbf{R}^n$ and scale matrix $\mathbf{C}$, a nxn positive definite matrix, if its density is (Dickey, 1967)

(1)

$$f(\mathbf{x}|\mathbf{m}, \mathbf{C}, D, d) = \frac{\Gamma[\frac{1}{2}(d+n)]D^{\frac{d}{2}}|\mathbf{C}|^{-\frac{1}{2}}}{\pi^{\frac{n}{2}}\Gamma[\frac{d}{2}]}\left\{D + (\mathbf{x}-\mathbf{m})'\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right\}^{-\frac{1}{2}(d+n)}, \quad \mathbf{x} \in \mathbf{R}^n,$$

where D and d are positive real numbers (typically, $D = d$ and d is the degrees of freedom or the kurthosis parameter). We say that $\mathbf{X} \sim t_n(\mathbf{m}, \mathbf{C}, D, d)$ or $\mathbf{X} \sim t_n(\mathbf{m}, \mathbf{C}, d)$, when $D = d$. It is not difficult to see that $\mathbf{X}$ has mean vector $\mathbf{m}$ for $d > 1$ and covariance matrix $\mathbf{\Sigma} = (d-2)^{-1}D\mathbf{C}$, for $d > 2$. Moreover, for $d > 2$, we can write the density in (1) in terms of $\mathbf{\Sigma}$ as follows

$$f(\mathbf{x}|\mathbf{m}, \mathbf{\Sigma}, d) = \frac{\Gamma[\frac{1}{2}(d+n)]|\mathbf{\Sigma}|^{-\frac{1}{2}}}{(\pi(d-2))^{\frac{n}{2}}\Gamma[\frac{d}{2}]}\left\{1 + \frac{(\mathbf{x}-\mathbf{m})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})}{d-2}\right\}^{-\frac{1}{2}(d+n)}, \quad \mathbf{x} \in \mathbf{R}^n.$$

When $\mathbf{m} = \mathbf{0}$ and $\mathbf{C} = \mathbf{I}_n$, the nxn identity matrix, the random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ is said to have a spherical t distribution (Kelker, 1970) and, provided they exist, we have that

$$E[X_{k+1}|X_k] = 0 \quad \text{and} \quad Var[X_{k+1}|X_k] = aX_k^2 + b, \tag{2}$$

where $a = (d-1)^{-1}$ and $b = (d-1)^{-1}D$.

It is well known that the $t_n(0, \mathbf{I}_n, D, d)$ is a mixture of the $N_n(0, v\mathbf{I}_n)$ in the scale parameter v. The mixing measure is the inverted gamma distributin with scale and shape parameters $D/2$ and $d/2$, respectively, wich we denote by $IG(D/2; d/2)$, with density given by

$$\pi(dv) = \frac{D^{\frac{d}{2}}}{2^{\frac{d}{2}}\Gamma[\frac{d}{2}]} v^{-\frac{1}{2}(d+2)} e^{-\frac{1}{2}Dv^{-1}} I_{(0,\infty)} dv, \tag{3}$$

where

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, it follows that a multivariate t distribution can be derived in two stages. In the first stage the conditional distribution of $\mathbf{X}$ given the scale parameter is specified. The second follows by specifying the prior distribution for the scale parameters. By adopting the predictivitic approach (de Finetti, 1931, 1937), this stages are replaced by the judgment about the observables (Iglesias, 1993; Wechsler, 1993). For example, the judgement of invariance under various groups of orthogonal transformation over infinite sequences of random quantities implies (Kingman, 1972; Smith, 1981; Diaconis, Eaton and Lauritzen, 1992) that the law of the sequence of observables can be represented as mixtures of conditionally independent normally distributed quantities. However, this type of condition does not permit one to characterize the mixing measure. Additional conditions can be examined to obtain the form of the mixing measure. Diaconis and Ylvisaker (1985) provided some results in this direction. They characterize a scale mixture of exponential distributions with mixing measure gamma by judgements of simmetry and additional conditions in terms of how one would predict X_{n+1} given $X_1, \ldots, X_n$. These results follows by using a theorem presented in Diaconis and Ylvisaker (1979), where the conjugate prior of a specified exponential family is characterized through the property of linear posterior expectations. This theorem is stated in Section 2. In Section 3, we use this result to characterize the mixing measure in the representation of infinite sequences of random quantities under special groups of orthogonal transformations. First, we consider orthogonally invariant random sequences and we show that if the observables satisfy the second property in (2) then any n-dimensional distribution is the spherical t distribution. Hence, we show that if the random sequence is invariant under orthogonal transformations that preserve the vector of ones and satisfy an additional condition similar to the second condition in (2) then, for each n, the n-dimensional distribution is the location mixture of the noncentral t distribution with the identity matrix as the scale matrix. Finally, in Section 4, the results of the previous sections are extended to cover the cases of orthogonally invariant (David, 1977; Diaconis, Eaton and Lauritzen, 1992) sequences of p-dimensional random vectors and it is shown that similar conditions implies that the mixing measure is the Wishart distribution.

2. Preliminaries

Diaconis and Ylvisaker (1979) provide a characterization of the conjugate priors for the natural parameter of a regular exponetial distribution. They show that linearity of the posterior expectations imply that the prior distribution must be conjugated. Their main result is stated in the sequel. Let μ be a σ-finite measure on the Borel sets of R^d. Consider the convex hull of the support set of the measure μ, and let $\mathcal{X}$ be the interior of this set. Assume that $\mathcal{X}$ is a nonempty open set in R^d. For $\theta \in \mathrm{R}^d$, define

$$M(\theta) = \log \int e^{x.\theta}\mu(dx)$$

and let $\Theta = \{\theta; M(\theta) < \infty\}$, a nonempty open set in R^d. The exponential family $\{P_\theta\}$ of probability measures through μ is determined by

$$P_\theta(dx) = e^{x.\theta - M(\theta)}\mu(dx), \quad \theta \in \Theta. \tag{4}$$

The parameter θ is called the natural parameter. Also, if X is a d-dimensional random vector with distribution P_θ of (4), then differentiating the identity

$$\int_{\mathcal{X}} P_\theta(dx) = 1,$$

with respect to θ and making adimissible interchanges of differentiation and integration, we find that

$$E[\mathrm{X}|\Theta] = M'(\Theta),$$

where

$$M'(\theta) = (\frac{\partial M(\theta)}{\partial\theta_1}, \ldots, \frac{\partial M(\theta)}{\theta_d})'.$$

Here Θ denotes a random vector and θ a particular value of Θ.

Theorem 1. *(Diaconis and Ylvisaker,1979). Suppose that Θ is open in R^d. Let X be a sample of size one from P_θ given in (4) and suppose that the support of μ contains an open interval in R^d. If Θ has a prior distribution, π, which does not concentrate at a single point and if*

$$E[M'(\Theta)|\mathrm{X}] = a\mathrm{X} + \mathrm{b},$$

for some constant a and a constant vector b, then $a \neq 0$, π is absolutely continuous with respect to the Lebesgue measure and has density

$$\pi(d\theta) = ke^{(a^{-1}b.\theta - a^{-1}(a-1)M(\theta))}d\theta.$$

3. Preditivistic characterization of the multivariate t distribution

Let O_n be the group of nxn orthogonal matrices. An infinite sequence of real random variables, say, $X_1, X_2, \ldots$, is said to be orthogonally invariant if, for each n, $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)'$ has an O_n-invariant distribution, that is, for each n, $\mathbf{X}^{(n)}$ and $\Gamma\mathbf{X}^{(n)}$ are identically distributed, for all $\Gamma \in O_n$. For example, if $Z_1, Z_2, \ldots$, have independent and identically distributed (i.i.d.) coordinates which are $N(0, \sigma^2)$, then the sequence is orthogonally invariant. Let P_σ denote the probability on $\mathbf{R}^\infty$ of $Z_1, Z_2, \ldots$, and π a probabilty measure on $[0, \infty)$. Hence, it is clear that

$$P_\pi = \int_{[0,\infty)} P_\sigma \pi(d\sigma)$$

is orthogonally invariant. It is also true that the law of all infinite sequences of orthogonally invariant random real variables can be represented in this maner. This representation has been shown to hold in a number of different contexts (Kelker, 1970; Kingman, 1972, Andrews and Mallows, 1974; Diaconis and Freedman, 1987). An interesting review on the subject can be found in Eaton (1989). We state the next result following Kingman (1972).

Theorem 2. *(Kingman, 1972) Let $X_1, X_2, \ldots$, be an infinite sequence of random variables which are orthogonally invariant. Then there is a non-negative real random variable V such that, conditional on V, the $X_1, X_2, \ldots$, are independent and normally distributed with mean zero and variance V.*

The random variable V defined in the previous theorem may be expressed more explicitly in a different number of ways. For instance, the strong law of large numbers shows at once that

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} X_i^2 = V,$$

with probability one. This fact is used in the next result.

Remark 1. *If in Theorem 2 we put the further condition that $X_1 = 0$ with probability zero then $\mu\{\sigma; \sigma > 0\} = 1$ (Eaton, 1981) where μ is the law of V. This fact follows from the identity*

$$P[X_1 = 0] = \int P_\sigma(X_1 = 0)\mu(d\sigma),$$

with $P_\sigma(X_1 = 0) = I_{\{0\}}(\sigma)$.

Proposition 1. *Let $X_1, X_2, \ldots$, be an infinite sequence of random variables orthoganally invariant such that $X_1 = 0$ with probability one. If $Var[X_2|X_1] = aX_1^2 + b$, for $0 < a < 1$ and $b > 0$, then, for any n, $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)'$ is distributed as $t_n(0, \mathbf{I}_n, a^{-1}b, a^{-1}(a+1))$. The converse also holds.*

Proof. By using Kingman's Theorem, the first condition implies that conditional on V, $\mathbf{X}^{(n)}$ is $N_n(0, V\mathbf{I}_n)$, where $V > 0$ with probability one.

Let $Y_i = X_i^2$, $i = 1, 2, \ldots$. Then, given $V = v$, the $Y_i's$ are iid gamma random variables with parameters $v/2$ and $1/2$, denoted by $\Gamma(v/2, 1/2)$ and setting $\theta = -(2v)^{-1}$, we have that

$$P_\theta(dy_1) = e^{\theta y_1 - M(\theta)}\mu(dy_1),$$

where

$$\mu(dy_1) = (\pi y_1)^{-\frac{1}{2}} I_{(0,\infty)}(y_1)dy_1$$

and

$$M(\theta) = -\frac{1}{2}\log(-\theta).$$

Using this fact and the second condition, standard properties of conditional expectations imply that

$$\begin{aligned} E[E\{Y_1|\Theta\}|Y_1] = E[E\{Y_1|V\}|Y_1] &= E[E\{Y_2|V\}|Y_1] \\ &= E[E\{Y_2|V,Y_1\}|Y_1] \\ &= E[Y_2|Y_1] \\ &= E[E\{Y_2|Y_1\}|Y_1^2] \\ &= aY_1 + b. \end{aligned}$$

By using Theorem 1, we obtain that

$$\pi(dv) = \frac{k}{2^{\frac{1}{2}a^{-1}(a+1)}} v^{-\frac{1}{2}(a^{-1}(a+1)+2)} e^{-\frac{1}{2}a^{-1}bv^{-1}} I_{(0,\infty)}dv.$$

Thus, it follows that V is distributed according to the $IG(a^{-1}b/2; a^{-1}(a+1)/2)$, which concludes the proof.

Proposition 1 implies that the law of the sequence $Y_1, Y_2, \ldots$, defined above can be represented as a scale mixture of the gamma distribution with the inverted gamma as a mixing measure. Moreover, for each n, $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ has an n-variate distribution with a density given by

$$f(\mathbf{y}) = \frac{\Gamma[\frac{1}{2}(a^{-1}(a+1)+n)]}{\pi^{\frac{n}{2}}\Gamma[\frac{1}{2}a^{-1}(a+1)]}(a^{-1}b)^{\frac{1}{2}a^{-1}(a+1)} \prod_{i=1}^{n} y_i^{\frac{1}{2}-1} \left\{a^{-1}b + \sum_{i=1}^{n} y_i\right\}^{-\frac{1}{2}[a^{-1}(a+1)+n]} I_{(0,\infty)^n}($$

where $(a,b)^n$ denotes the cartesian product of (a,b) n-folds. This distribution may be considered as a multivariate version of the ordinary *BetaII* distribution (Olkin and Rubin, 1964; Tan, 1969) and is called the n-variate inverted Dirichlet in Tiao and Guttman (1965), where it is denoted by

$$a^{-1}bID_n(\frac{1}{2}, \ldots, \frac{1}{2}; \frac{1}{2}a^{-1}(a+1)).$$

The notation and the result that follows next are used to characterize a class of orthogonally invariant random variables which can be represented as a location mixture of a multivariate noncentral t distribution. Let $O_n(\mathbf{1}_n) = \{\mathbf{A} \in O_n; \mathbf{A}\mathbf{1}_n = \mathbf{1}_n\}$, where $\mathbf{1}_n$ is a vector of ones of dimension n.

Theorem 3. *(Smith, 1981) Let $X_1, X_2, \ldots,$ be an infinite sequence of real random variables such that, for any n, the distribution of $\mathbf{X}^{(n)}$ is $\mathcal{O}_n(\mathbf{1}_n)$ invariant. Then, there are real random variables M and (nonnegative) V, such that conditional on M and V, $X_1, X_2, \ldots,$ are independent and $N(M, V)$.*

As emphasized by Smith (1981), this result is of particular interest to Bayesian statisticians because it provides a characterization of the form of the subjective assesment about the sequence $\mathbf{X}^{(n)}$ which would justify using a normal likelihood together with a prior distribution for (M, V). The parameters M and V may be interpreted as the limits of the sequences $\{\bar{X}_n\}$ and $\{S_n^2\}$, that is,

$$\lim_{n\to\infty} \bar{X}_n = M, \quad \text{and} \quad \lim_{n\to\infty} S_n^2 = V,$$

with probability one, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i/n$ and $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Remark 2. *Using similar arguments as in in Remark 1, it can be shown that if in Theorem 3 we include the additional condition that $X_1 = X_2$ with probability zero, then $V > 0$ with probability one.*

Proposition 2. *Let $X_1, X_2, \ldots,$ be an infinite sequence of real random variables such that, for any n, the distribution of $\mathbf{X}^{(n)}$ is a $\mathcal{O}_n(\mathbf{1}_n)$ invariant distribution. If, additionally,*

$$E[(X_2 - M)^2 | X_1, M] = a(X_1 - M)^2 + b, \tag{5}$$

for $0 < a < 1$ and $b > 0$, then the law of $\mathbf{X}^{(n)}$ is a location mixture of $t_n(M, \mathbf{I}_n, a^{-1}b, a^{-1}(a+1))$. Furthermore, M and V are independent.

Proof. Standard properties of conditional expectations and assumption (5) imply that

$$\begin{aligned}
E[E\{(X_1 - M)^2 | M, V\} | (X_1 - M)^2, M] &= E[E\{(X_2 - M)^2 | M, V\} | (X_1 - M)^2, M] \\
&= E[E\{(X_2 - M)^2 | (X_1 - M)^2, M, V\} | (X_1 - M)^2, M] \\
&= E[(X_1 - M)^2 | (X_1 - M)^2, M] \\
&= E[E\{(X_1 - M)^2 | X_1, M)\} | (X_1 - M)^2, M] \\
&= a(X_1 - M)^2 + b.
\end{aligned}$$

Let P_m be the conditional law of $X_1, X_2, \ldots,$ given $M = m$ and let E_m be the expectation operator with respect to P_m. Then, by the previous identities we have that

$$E_m[E_m\{(X_1 - m)^2 | V\} | (X_1 - m)^2] = a(X_1 - m)^2 + b.$$

Now, if $P_m^{(n)}(.|v)$ denote the P_m-law of $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)'$, we have that $P_m^{(n)}(.|v)$ is the $N_n(m\mathbf{1}_n, v\mathbf{I}_n)$. Moreover, since for each real m, the sequence $X_1 - m, X_2 - m, \ldots,$ is orthogonally invariant relative to P_m, it follows from Proposition 1 that

$$\pi_m(dv) = kv^{-\frac{1}{2}(a^{-1}(a+1)+2)} e^{-\frac{1}{2}a^{-1}bv^{-1}} I_{(0,\infty)}(v), \tag{6}$$

which implies that

$$\int P_m^{(n)}(.|v)\pi_m(dv)$$

is the $t_n(m, I_n, a^{-1}b, a^{-1}(a+1))$. Finally, we note that the conditional distribution of V given $M = m$ does not depend on m. Hence M and V are independent, which concludes the proof.

4. A Predictivistic characterization of the matric-variate t distribution

Here the results of the previous section are extended to the matrix case. In the sequel, $A \otimes B$ denotes the Kronecker product of the mxm matrix A and the pxq matrix B which is defined as the $mpxnq$ matrix with entries

$$(A \otimes B)_{ir,js} = a_{ij}b_{rs},$$

where $A = \{a_{ij}\}$ and $B = \{b_{rs}\}$.

The nxp random matrix X has a matrix-variate t distribution, say $X \sim t(M, C, D; d)$, where M is a nxp location matrix and C and D are nxn and pxp positive definite matrices and $d + (n-1)(p-1) > 0$, if the density of X is given by (Dickey, 1967)

$$f(X|M, C, D, d) = \frac{\Gamma_p[\frac{1}{2}(d+np)]|D|^{\frac{1}{2}(d+(n-1)p)}|C|^{-\frac{1}{2}p}}{\pi^{\frac{1}{2}np}\Gamma_p[\frac{1}{2}(d+n(p-1))]}|D+(X-M)'C^{-1}(X-M)|^{-\frac{1}{2}(d+np)},$$

where

$$\Gamma_p[\frac{1}{2}q] = \pi^{\frac{1}{4}p(p-1)}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(q-i+1)].$$

The above notation differs from that of Dickey (1967) and Dawid (1981), who denote the above distribution by $t(C, D, M, d+np)$ and $t(d+(n-1)(p-1); M, C, D)$, respectively. The major diference being in the degrees of fredom parameter. It is also known that the $t(0, I_n, D, d)$ is a mixture of $N(0, I_n \otimes V)$, in the pxp positive definite scale matrix V. The mixing measure is the Inverted Wishart with parameters D and $d + n(p-1)$, wich we denote by $IW_p(D, d + n(p-1))$, with density given by

$$\pi(dV) = \frac{|D|^{\frac{1}{2}(d+n(p-1))}}{2^{\frac{1}{2}(d+n(p-1))}\Gamma_p[\frac{1}{2}(d+n(p-1))]}|V|^{-\frac{1}{2}(d+n(p-1)+p+1)}e^{tr[-\frac{1}{2}DV^{-1}]}dV,$$

for any positive definite matrix V and zero otherwise. By using similar arguments as the ones considered in the previous section, we characterize this mixing measure by setting additional conditions on the orthogonally invariant random column vectors of the matrix X. First we state a theorem due to Dawid (1978). Let $X_1, X_2, \ldots,$ be a sequence of p-dimensional random column vectors. For each k, let $X^{(k)}$ be the kxp matrix whose rows are the transposed vectors $X_1', \ldots, X_k'$. The group of kxk orthogonal matrices, O_k, acts on $X^{(k)}$ by left multiplication. The sequence $X_1, X_2, \ldots,$ is said to be orthogonally invariant if for each k, the law of $\mathbf{X}^{(k)}$, say $P^{(k)}$, is $\mathcal{O}_k$ invariant. Now, let P be the law of the sequence $\mathbf{X}_1, \mathbf{X}_2, \ldots$, and let Q_Σ denote the law of $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ when the $\mathbf{Y}_i'$s are independent and identically distributed $N_p(0, \Sigma)$ and $Q_{n\Sigma}$ is the $N(0, \mathbf{I}_n \otimes \Sigma)$ distribution.

Theorem 4. *(Dawid, 1978). If $\mathbf{X}_1, \mathbf{X}_2, \ldots$, is orthogonally invariant, then there is a unique probability measure μ defined on the Borel sets of $\mathbf{S}_p^+$, the set of all pxp symetric nonnegative definite matrices, such that*

$$P = \int Q_\sigma \mu(d\Sigma).$$

In particular, for each k,

$$P^{(k)} = \int Q_{k\Sigma} \mu(d\Sigma).$$

The converse is also true.

In other words, if $\mathbf{X}_1, \mathbf{X}_2, \ldots$, is orthogonally invariant, then there exists a *pxp* symmetric nonnegative definite matrix $\mathbf{V}$, such that for each k, conditionally on $\mathbf{V}$, $\mathbf{X}_1, \ldots, \mathbf{X}_k$ are independent and identically distributed $N_p(0, \mathbf{V})$. From the strong law of the large numbers, it follows that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i' = \mathbf{V},$$

with probability one.

Remark 3. *If If in Theorem 4 we include the additional condition that $\mathbf{X}_1, \ldots, \mathbf{X}_p$ are linerly independent with probability zero, then*

$$\mu\{\Sigma; \quad \Sigma \quad \text{is} \quad \text{positive} \quad \text{definite}\} = 1,$$

where μ is the law of $\mathbf{V}$. This follows similarly as the result in Remark 1 and means that if $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are random identically distributed $N_n(0, \Sigma)$ vectors with Σ positive definite, then $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are linearly dependent with probability zero (Muirhead, 1982).

Proposition 3. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$, be an infinite sequence of p-dimensional column vectors which are orthogonally invariant and for k fixed let $\mathbf{X}_i^{(k)} = (\mathbf{X}_{(i-1)k+1}, \ldots, \mathbf{X}_{ik})'$, $i = 1, 2, \ldots$. If $\mathbf{X}_1, \ldots, \mathbf{X}_p$ are linearly independent with probability one and*

$$E[\mathbf{X}_2^{(p)'} \mathbf{X}_2^{(p)} | \mathbf{X}_1^{(p)}] = a \mathbf{X}_1^{(p)'} \mathbf{X}_1^{(p)} + \mathbf{B}, \tag{8}$$

where $0 < a < 1$ and $\mathbf{B}$ is a pxp positive definite matrix, then the distribution of $\mathbf{X}^{(n)}$ is the $t(0, \mathbf{I}_n, a^{-1}\mathbf{B}, a^{-1}(a+p) - n(p-1))$distribution.

Proof. The first condition and the previous theorem implies that, for each n, the distribution of $\mathbf{X}^{(n)}$ given $\mathbf{V}$ is $N(0, \mathbf{I}_n \otimes \mathbf{V})$, where $\mathbf{V}$ is a positive definite matrix with probability one. Let $\mathbf{Y}_i' = \mathbf{X}_i^{(p)'} \mathbf{X}_i^{(p)}$, $i = 1, 2, \ldots$. Because $\mathbf{V}$ is positive definite then, given $\mathbf{V}$, the

$\mathbf{Y}_i's$ are pxp positive definite matrices with probability one (Remark 3) and identically distributed according to the p-dimensional Wishart distribution with parameters $\mathbf{V}$ and p, which we denote by $W_p(\mathbf{V},p)$. Setting $\Theta = -(2\mathbf{V})^{-1}$ we obtain a $\frac{1}{2}p(p+1)$-variate density for $\mathbf{Y}_1$, which we write as

$$P_\Theta(dY_1) = e^{tr[\Theta Y_1] + \frac{1}{2}p\log|-\Theta|}\mu(dY_1),$$

where

$$\mu(dY_1) = \{\Gamma[\frac{1}{2}p]\}^{-1}|\mathbf{Y}_1|^{-1/2}dY_1.$$

Standard properties of conditional expectations and (7) imply that

$$\begin{aligned} E[E\{\mathbf{Y}_1|\Theta\}|\mathbf{Y}_1] &= E[E\{\mathbf{Y}_1|\mathbf{V}\}|\mathbf{Y}_1] \\ &= E[E\{\mathbf{Y}_2|\mathbf{V}\}|\mathbf{Y}_1] \\ &= E[E\{\mathbf{Y}_2|\mathbf{Y}_1,\mathbf{V}\}|\mathbf{Y}_1] \\ &= E[\mathbf{Y}_2|\mathbf{Y}_1] \\ &= a\mathbf{Y}_1 + \mathbf{B}. \end{aligned}$$

Now, using Theorem 1, it follows that

$$\pi(d\Theta) = k|-\Theta|^{\frac{1}{2}pa^{-1}(a-1)}e^{tr[a^{-1}\mathbf{B}\Theta]}d\Theta,$$

which implies that $-\Theta$ is distributed according to the $W_p(2a^{-1}\mathbf{B}, a^{-1}(a+p))$. Consequently, $\mathbf{V}$ is distributed according to the $IW_p(a^{-1}\mathbf{B}, a^{-1}(a+p))$, which proves the theorem.

An extension of Proposition 2 follows by making minor changes in Proposition 3 and using some results in Diaconis et al. (1992). Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be p-dimensional random column vectors, forming the nxp matrix

$$\mathbf{X}^{(n)} = \begin{pmatrix} \mathbf{X}_1' \\ \vdots \\ \mathbf{X}_n' \end{pmatrix}.$$

The group $\mathbf{O}_n(\mathbf{1}_n)$ acts on $\mathbf{X}^{(n)}$ by left multiplication. The law of $\mathbf{X}^{(n)}$ is asummed to be $O_n(\mathbf{1}_n)$-invariant. Let $P^{(k)}$ be the law of $\mathbf{X}^{(k)}$ and

$$Q_\mu^{(k)} = \int N(\mathbf{1}_k\mathbf{m}', \mathbf{I}_k\otimes\Sigma)\mu(dm, d\Sigma),$$

where $\mathbf{m} = (m_1,\ldots,m_p)'$ is in $\mathbf{R}^p$.

Theorem 5. *(Diaconis, Eaton and Lauritzen, 1992). Suppose that there exists a probability measure μ on $R^p x S_p^+$, such that for all k with $p+k+2 < n-1$, the variation distance between $P^{(k)}$ and $Q^{(k)}$ is bounded above by*

$$2\{(1-\frac{k+p+2}{n-1})^{-c}-1\}+2\{(\frac{n}{n-2})^{\frac{1}{2}p}-1\},$$

where $c = t^2/2$ and $t = min\{k,p\}$.

Let now $\mathbf{X}_1, \mathbf{X}_2, \ldots$, an infinite sequence of p-dimensional random column vectors such that for each n, the law of $\mathbf{X}^{(n)}$, say, $P^{(n)}$, is $O_n(1_n)$-invariant and let P be the law of the infinite sequence. Suppose also that $E[\mathbf{X}_1] < \infty$ and $E[\mathbf{X}_1\mathbf{X}_1'] < \infty$. Hence, with this condition, an infinite version of Theorem 5 can be obtained by using similar arguments as in Diaconis et al. (1992). In fact, from the infinite exchangeability of $\{\mathbf{X}_n; n \in N\}$, if $E[\mathbf{X}_1] < \infty$ and $E[\mathbf{X}_1\mathbf{X}_1'] < \infty$, then the there exists a σ-algebra $\mathcal{T}$ of events such that (Chow and Taicher, 1978)

$$\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \stackrel{a.s.}{\to} E[\mathbf{X}_1|\mathcal{T}] = \mathbf{M},$$

and

$$\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{X}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)' \stackrel{a.s.}{\to} E[\mathbf{X}_1\mathbf{X}_1'|\mathcal{T}] - E[\mathbf{X}_1|\mathcal{T}]\{E[\mathbf{X}_1|\mathcal{T}]\}' = \mathbf{V}.$$

By stting μ_n as the P-law of $(\bar{\mathbf{X}}_n, \mathbf{S}_n)$ and using Theorem 5 it follows that $P^{(k)} = Q_\mu^{(k)}$, for all k, where μ is the measure of $(\mathbf{M}, \mathbf{V})$.

Let, now

$$\mathbf{X}_k^{(p)} = \begin{pmatrix} \mathbf{X}_{(k-1)p} \\ \vdots \\ \mathbf{X}_{kp}' \end{pmatrix},$$

and $\mathbf{Z}^{(p)} = \mathbf{X}_1^{(p)} - \mathbf{X}_2^{(p)}$. It follows by virtue of arguments similar to the ones considered in Remark 3 that if

$$P[\mathbf{Z}_1, \ldots, \mathbf{Z}_p \text{are} \quad \text{linearly} \quad \text{dependent}] = 0,$$

where $\mathbf{Z}_i = \mathbf{X}_i - \mathbf{X}_{p+1}$, $i = 1, \ldots, p$, then

$$\mu\{\mathbf{R}^p x\{\mathbf{V}; \mathbf{V} \quad \text{is} \quad \text{positive} \quad \text{definite}\}\} = 1.$$

Proposition 4. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$, be an infinite sequence of p-dimensional random column vectors such that for all n, $\mathbf{X}^{(n)}$ is $O_n(1_n)$-invariant. If $\mathbf{M}$, $\mathbf{V}$ and $\mathbf{Z}_p$ are as defined above and if*

$$E[(\mathbf{X}_2^{(p)} - \mathbf{1}_p\mathbf{M}')'(\mathbf{X}_2^{(p)} - \mathbf{1}_p\mathbf{M}')|\mathbf{X}_i^{(p)}, \mathbf{M}] = a(\mathbf{X}_1^{(p)} - \mathbf{1}_p\mathbf{M}')'(\mathbf{X}_1^{(p)} - \mathbf{1}_p\mathbf{M}') + \mathbf{B},$$

where $0 < a < 1$ *and* B *is a pxp symmetric positive definite matrix, then* $\mathbf{X}^{(n)}$ *is a mixture in the location parameter of* $t(\mathbf{1}_n'\mathbf{M}, \mathbf{I}_n, a^{-1}\mathbf{B}, a^{-1}(a+p)-n(p-1))$. *Furthermore,* M *and* V *are independent.*

Proof. Using similar arguments as in the proof of Proposition 2 we have

$$E[E\{(\mathbf{X}_1-\mathbf{1}\mathbf{M}')'(\mathbf{X}_1^{(p)}-\mathbf{1}_p\mathbf{M}')|\mathbf{M},\mathbf{V}\}|\mathbf{X}_1^{(p)},\mathbf{M}] = E[(\mathbf{X}_2^{(p)}-\mathbf{1}_p\mathbf{M}')'(\mathbf{X}_2^{(p)}-\mathbf{1}_p\mathbf{M}')|\mathbf{X}_1^{(p)},\mathbf{M}]$$

$$= a(\mathbf{X}_1^{(p)}-\mathbf{1}_p\mathbf{M}')'(\mathbf{X}_1-\mathbf{1}_p\mathbf{M}')+\mathbf{B}. \tag{8}$$

Let P_m be a regular conditional probability version of $\mathbf{X}_1, \mathbf{X}_2, \ldots$ given $\mathbf{M} = \mathbf{m}$ and $E_m[\cdot]$ the expectation operator relative to P_m. Then the P_m-law of $\mathbf{Y}_1 = \mathbf{X}_1 - \mathbf{1}_p\mathbf{m}'$ given V is $N(\mathbf{1}_p\mathbf{m}', \mathbf{I}_p \otimes \mathbf{V})$. Morerover, (8) implies that

$$E_m[E_m\{\mathbf{Y}_1\mathbf{Y}_1'|\mathbf{V}\}|\mathbf{Y}_1] = a\mathbf{Y}_1 + \mathbf{B},$$

which together with Proposition 3 concludes the proof.

References

Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc.*, B, 36, 99-101.

Chow, Y.S. and Teicher, H. (1978). *Probability Theory.* Springer-Verlag. New York

Dawid, A.P. (1978). Extendability of spherical matrix distributions. *J. Multiv. Anal.*, 8, 559-566.

Dawid, A.P. (1981). Some matrix-variate distribution theory; notational considertaion and a Bayesian application. *Biometrika*, 68, 1, 265-274.

de Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Memorie della R. Academia Nazionale dei Lincei*, 4, 86-133.

de Finetti, B. (1938). Sur la condition d'equivalence partielle. *Actualitès Scientifique et Industrielles*, No. 739. Hermon and Cii. Paris. Translated in *Studies in Inductive Logic and Probability*, II, University of California Press, Berkeley (ed. R. Jeffrey).

de Finetti, B. (1972). *Probability, Induction and Statistics.* Wiley. New York.

Diaconis, P. and Freedman, D. (1987). A dozen de Finetty-style results in search of a Theory. *Ann. Inst. Henri Poincaré, Probabilités et Statistique.* 23, 397-423.

Diaconis, E., Eaton, M. and Lauritzen, S. (1992). Finite de Finetti theorems in linear models and multivariate analysis. *Scand. J. Statist.*, 19, 289-315.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Ann. of Statistic.*, 7, 269-281.

Diaconis, P. and Yvisaker, D. (1985). Quantifying prior opinion. In Bayesian Statistics 2. Elsevier Science Publisher B.V. and Valencia University Press.

Dickey, J.M. (1967). Matric-variate generalization of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.*, 38, 511-518.

Eaton, M. (1981). On the projections of isotropic distributions. *Ann. Statist.*, 9, 391-400.

Eaton, M. (1989). Group invariance; applications in statistics. Vol. 1 in *Regional Conference Series in Probability and Statistics.* Institute of Mathematical Statistics and American Statistical Association. Hayward, CA.

Iglesias, P. (1993). Finite forms of the de Finetti's Theorem: A predictivistic approach to statistical inference in finite populations (In Portuguese). Doctoral Thesis. Instituto de Matemática e Estatística, Universidade de São Paulo.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Shankhya,* A, 32, 419-430.

Kingman, J.F.C. (1972). On random sequences with spherical symmetry. *Biometrika*, 59, 183-197.

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

Olkin, I. and Rubin, H. (1964). Multivariate beta distribution and independence properties of the Wishart distribution. *Ann. Math. Statistic.*, 35, 261-269.

Smith, A.F.M. (1981). On random sequences with centered spherical symmetry. *J. Roy. Statist. Soc.*, B, 43, 203-209.

Tan, W.Y. (1969). Note on the multivariate and generalized multivariate beta distributions. *J. Am. Statist. Assoc.*, 64, 230-241.

Tiao, G.C. and Guttman, I. (1965). The multivariate inverted beta distribution with applications. *J. Am. Statist. Assoc.*, 60, 793-805.

Wechsler, S. (1993). Exchangeability and predictivism. *Erkenntnis - international Journal of Analytic Philosophy*, 38, 343-350.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1993

9301 - BUENO, V.C. & ARIZONO, H. Comparisons for Maintenance Policies Involving Complete and Minimal Repair Through Compensator Transform, 12p.

9302 - BUENO, V.C. Maintenance Comparisons Through Compensator Transform: Block Policies, 13p.

9303 - FERRARI, P.A. & FONTES, L.R.G. Shock Fluctuations in the Asymmetric Simple Exclusion Process, 17p.

9304 - CORDEIRO, G. M., BOTTER, D.A. & FERRARI, S.L.P. Nonnull Asymptotic Distributions of Three Classic Criteria in Generalised Linear Models, 21p.

9305 - TANAKA, N.I. & TAVARES, H.R. Introdução à Teoria da Percolação, 38p.

9306 - BOLFARINE, H., ZACKS, S. & SANDOVAL, M.C. On Predicting the Population Total Under Regression Models with Measurement Erros, 10p.

9307 - FERRARI, P.A. & FONTES, L.R.G. The Net Output Process of a System with Infinitely Many Queues, 20p.

9308 - HO, L.L. & BUSSAB, W.O. Decomposição de Dados Categóricos com Auxílio de Árvores Dicotômicas, 23p.

9309 - BOLFARINE, H. & VALLE, R.B.A. Robust Modelling in Measurement Error Models Using the t Distribution, 14p.

9310 - FERRARI, S.L. de P. & ARELLANO, R.B. Bartlett Corrected Test for Regression Models with Student-t Independent Errors, 30p.

9311 - PAULINO, C.D.M. & PEREIRA, C.A. de B. Bayesian Methods for Categorical Data Under Informative General Censoring, 13p.

9312 - INOUE, L.Y.T., PASCHOA, O.E.D. & WECHSLER, S. Bayesian Assessment of Bioequivalence, 9p.

9313 - MADRUGA, M.R., PEREIRA, C.A. de B. & RABELLO-GAY, M.N. Bayesian Dosimetry: Radiation Dose Versus Frequencies of Cells with Aberrations, 15p.

9314 - MENTZ, R.P., MORETTIN, P.A. & TOLOI, C.M.C. On Residual Variance Estimation in Arma Models, 47p.

9315 - PAULINO, C.D.M. & PEREIRA, C.A. de B. On Identifiability of Parametric Statistical Models, 35p.

9316 - BARLOW, R. E. & PEREIRA, C.A. de B. The Bayesian Operation and Probabilistic Influence Diagrams, 50p.

9317 - PAULA, G.A. & CORDEIRO, G.M. Bias Correction and Improved Residuals for Non-Exponential Family Nonlinear Models, 18p.

9318 - IGLESIAS, P.L.Z., PEREIRA, C.A. de B. & TANAKA, N.I. Finite Forms of de Finetti's Type Theorem for Univariate Uniform Distributions, 25p.

9319 - BOLFARINE, H., NASCIMENTO, J.A. & ROJAS, M.G. Maximum Likelihood Estimation of Linear Structural Relationships Via the EM Algorithm, 13p.

9320 - BOLFARINE, H. & ROJAS, M.G. Structural Comparative Calibration Using the EM Algorithm, 17p.

9321 - RODRIGUES, J. & ELIAN, S.N. Maximum Likelihood Prediction in Finite Populations, 13p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
IME-USP
Caixa Postal 20.570
01452-990 - São Paulo, Brasil