Multivariate t Distributions and Their Applications

Samuel Kotz
George Washington University

Saralees Nadarajah
University of South Florida

CAMBRIDGE
UNIVERSITY PRESS

'HE UNIVERSITY OF CAMBRIDGE
ridge, United Kingdom

J, UK
.1, USA
: 3207, Australia

l, South Africa

exception
nsing agreements,
ithout
y Press.

em LaTeX 2_ε [AU]

the British Library.

Data

ons / Samuel Kotz,

ability theory) I. Nadarajah,

Contents

Preface

Multivariate t distributions have attracted somewhat limited attention of researchers for the last 70 years in spite of their increasing importance in classical as well as in Bayesian statistical modeling. These distributions have been perhaps unjustly overshadowed – during all these years – by the multivariate normal distribution. Both the multivariate t and the multivariate normal are members of the general family of elliptically symmetric distributions. However, we feel that it is desirable to focus on these distributions separately for several reasons:

- Multivariate t distributions are generalizations of the classical univariate Student t distribution, which is of central importance in statistical inference. The possible structures are numerous, and each one possesses special characteristics as far as potential and current applications are concerned.
- Application of multivariate t distributions is a very promising approach in multivariate analysis. Classical multivariate analysis is soundly and rigidly tilted toward the multivariate normal distribution while multivariate t distributions offer a more viable alternative with respect to real-world data, particularly because its tails are more realistic. We have seen recently some unexpected applications in novel areas such as cluster analysis, discriminant analysis, multiple regression, robust projection indices, and missing data imputation.
- Multivariate t distributions for the past 20 to 30 years have played a crucial role in Bayesian analysis of multivariate data. They serve by now as the most popular prior distribution (because elicitation of prior information in various physical, engineering, and financial phenomena is closely associated with multivariate t distributions) and generate meaningful posterior distributions. This diversity and the apparent

ease of applications require careful analysis of the properties of the distribution in order to avoid pitfalls and misrepresentation.

The compilation of this book was a somewhat daunting task (as our Contents indicates). Indeed, the scope of the multivariate t distributions is unsurpassed, and, although there are books dealing with multivariate continuous distributions and review articles in the *Encyclopedia of Statistical Sciences and Biostatistics*, the material presented in these sources is quite limited.

Our goal was to collect and present in an organized and user-friendly manner all of the relevant information available in the literature worthy of publication. It is our hope that the readers – both novices and experts – will find the book useful. Our thanks are due to numerous authors who generously supplied us with their contributions and to Lauren Cowles, Elise Oranges and Lara Zoble at Cambridge University Press for their guidance. We also wish to thank Anusha Thiyagarajah for help with editing.

Samuel Kotz
Saralees Nadarajah

lysis of the properties of the
d misrepresentation.

ewhat daunting task (as our
the multivariate t distribu-
ire books dealing with multi-
articles in the *Encyclopedia*
e material presented in these

organized and user-friendly
able in the literature worthy
rs – both novices and experts
ue to numerous authors who
ions and to Lauren Cowles,
e University Press for their
Thiyagarajah for help with

Samuel Kotz
Saralees Nadarajah

1
Introduction

1.1 Definition

There exist quite a few forms of multivariate t distributions, which will be discussed in subsequent chapters. In this chapter, however, we shall describe the most common and natural form. It directly generalizes the univariate Student's t distribution in the same manner that the multivariate normal distribution generalizes the univariate normal distribution.

A p-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_p)^T$ is said to have the p-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$ (and with $\boldsymbol{\Sigma}$ denoting the corresponding covariance matrix) if its joint probability density function (pdf) is given by

$$f(\mathbf{x}) = \frac{\Gamma((\nu+p)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)\left|\mathbf{R}\right|^{1/2}}\left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{R}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-(\nu+p)/2}. \tag{1.1}$$

The degrees of freedom parameter ν is also referred to as the shape parameter, because the peakedness of (1.1) may be diminished, preserved, or increased by varying ν (see Section 1.4). The distribution is said to be central if $\boldsymbol{\mu} = \mathbf{0}$; otherwise, it is said to be noncentral.

Note that if $p = 1$, $\boldsymbol{\mu} = 0$, and $\mathbf{R} = 1$, then (1.1) is the pdf of the univariate Student's t distribution with degrees of freedom ν. These univariate marginals have increasingly heavy tails as ν decreases toward unity. With or without moments, the marginals become successively less peaked about $0 \in \Re$ as $\nu \downarrow 1$.

If $p = 2$, then (1.1) is a slight modification of the bivariate surface of Pearson (1923). If $\nu = 1$, then (1.1) is the p-variate Cauchy distribution. If $(\nu+p)/2 = m$, an integer, then (1.1) is the p-variate Pearson type VII distribution. The limiting form of (1.1) as $\nu \to \infty$ is the joint pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Hence, (1.1) can be viewed as an approximation of the multivariate normal distribution. The particular case of (1.1) for $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{R} = \mathbf{I}_p$ is a mixture of the normal density with zero means and covariance matrix $v\mathbf{I}_p$ – in the scale parameter v. The class of elliptically contoured distributions (see, for example, Fang et al., 1990) contain (1.1) as a particular case. Also (1.1) has the attractive property of being Schur-concave when elements of $\mathbf{R}$ satisfy $r_{ij} = \rho$, $i \neq j$ (see Marshall and Olkin, 1974). Namely, if $\mathbf{a}$ and $\mathbf{b}$ are two p-variate vectors with components ordered to achieve $a_1 \geq a_2 \geq \cdots \geq a_p$ and $b_1 \geq b_2 \geq \cdots \geq b_p$, and if this ordering implies $\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i$ for $k = 1, 2, \ldots, p-1$ and $\sum_{i=1}^{p} a_i \leq \sum_{i=1}^{p} b_i$, then (1.1) satisfies $f(\mathbf{a}) \geq f(\mathbf{b})$.

In Bayesian analyses, (1.1) arises as: (1) the posterior distribution of the mean of a multivariate normal distribution (Geisser and Cornfield, 1963; see also Stone, 1964); (2) the marginal posterior distribution of the regression coefficient vector of the traditional multivariate regression model (Tiao and Zellner, 1964); (3) the marginal prior distribution of the mean of a multinormal process (Ando and Kaufman, 1965); (4) the marginal posterior distribution of the mean and the predictive distribution of a future observation of the multivariate normal structural model (Fraser and Haq, 1969); (5) an approximation to posterior distributions arising in location-scale regression models (Sweeting, 1984, 1987); and (6) the prior distribution for set estimation of a multivariate normal mean (DasGupta et al., 1995). Additional applications of (1.1) can be seen in the numerous books dealing with the Bayesian aspects of multivariate analysis.

1.2 Representations

If $\mathbf{X}$ has the p-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$, then it can be represented as

- If $\mathbf{Y}$ is a p-variate normal random vector with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, and if $\nu S^2/\sigma^2$ is the chi-squared random variable with degrees of freedom ν, independent of $\mathbf{Y}$, then

$$\mathbf{X} = S^{-1}\mathbf{Y} + \boldsymbol{\mu}. \tag{1.2}$$

  This implies that $\mathbf{X} \mid S = s$ has the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $(1/s^2)\boldsymbol{\Sigma}$.

1) as $\nu \to \infty$ is the joint pdf of the
an vector $\boldsymbol{\mu}$ and covariance matrix
approximation of the multivariate
ase of (1.1) for $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{R} = \mathbf{I}_p$
h zero means and covariance ma-
The class of elliptically contoured
et al., 1990) contain (1.1) as a
tractive property of being Schur-
$_{ij} = \rho$, $i \neq j$ (see Marshall and
wo p-variate vectors with compo-
$\geq a_p$ and $b_1 \geq b_2 \geq \cdots \geq b_p$, and
$_{=1} b_i$ for $k = 1, 2, \ldots, p-1$ and
$f(\mathbf{a}) \geq f(\mathbf{b})$.
(1) the posterior distribution of
tribution (Geisser and Cornfield,
arginal posterior distribution of
traditional multivariate regres-
) the marginal prior distribution
(Ando and Kaufman, 1965); (4)
he mean and the predictive dis-
multivariate normal structural
approximation to posterior dis-
ession models (Sweeting, 1984,
set estimation of a multivariate
Additional applications of (1.1)
ing with the Bayesian aspects of

ations

th degrees of freedom ν, mean
it can be represented as

or with mean $\mathbf{0}$ and covariance
squared random variable with
Y, then

$+ \boldsymbol{\mu}$. (1.2)

ariate normal distribution with
$(1/s^2)\boldsymbol{\Sigma}$.

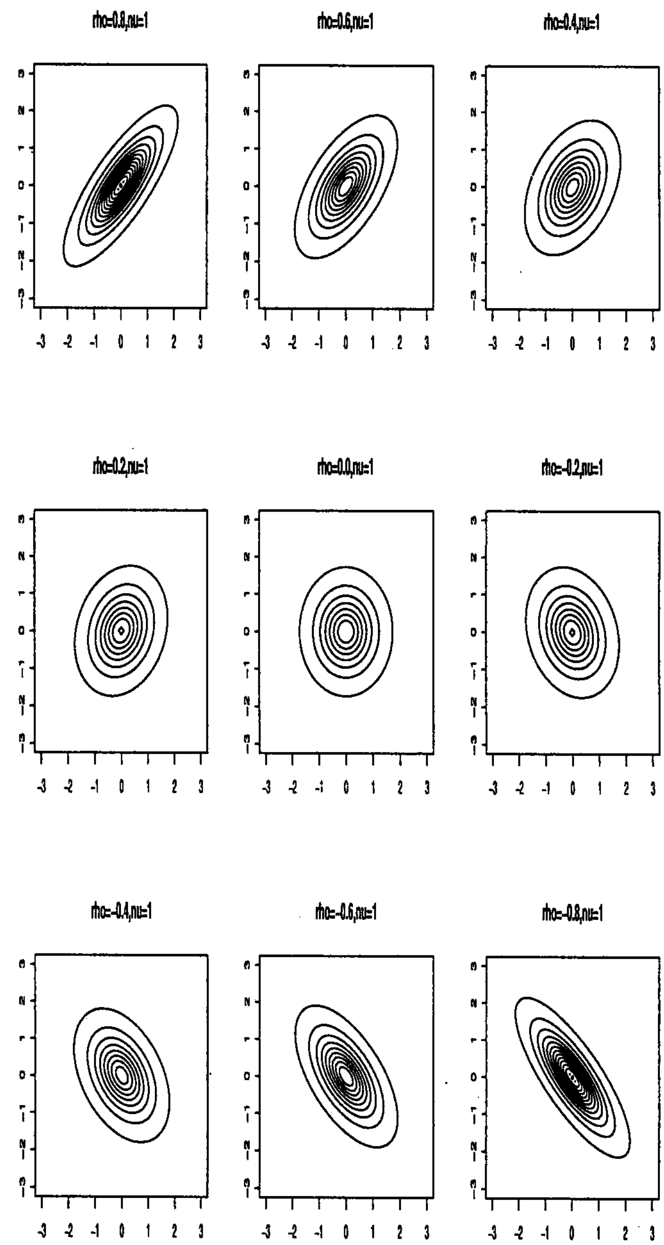

Fig. 1.1. Joint contours of (1.1) with degrees of freedom $\nu = 1$, zero means, and correlation coefficient $\rho = 0.8, 0.6, \ldots, -0.6, -0.8$

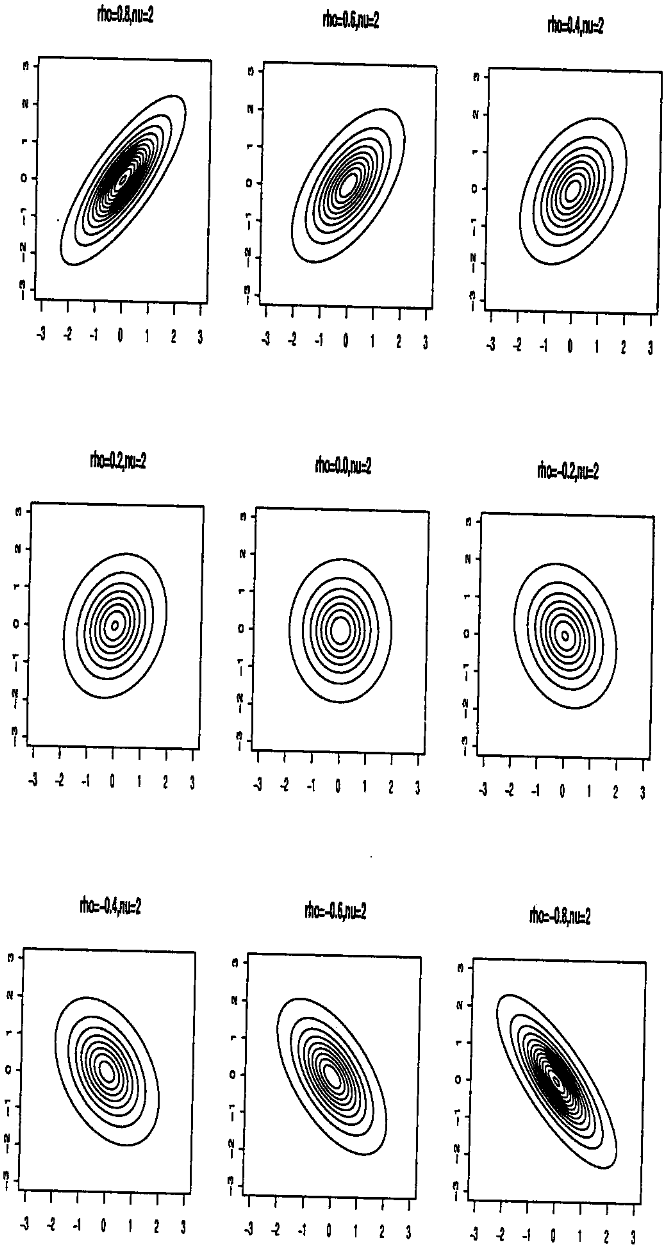

Fig. 1.2. Joint contours of (1.1) with degrees of freedom $\nu = 2$, zero means, and correlation coefficient $\rho = 0.8, 0.6, \ldots, -0.6, -0.8$

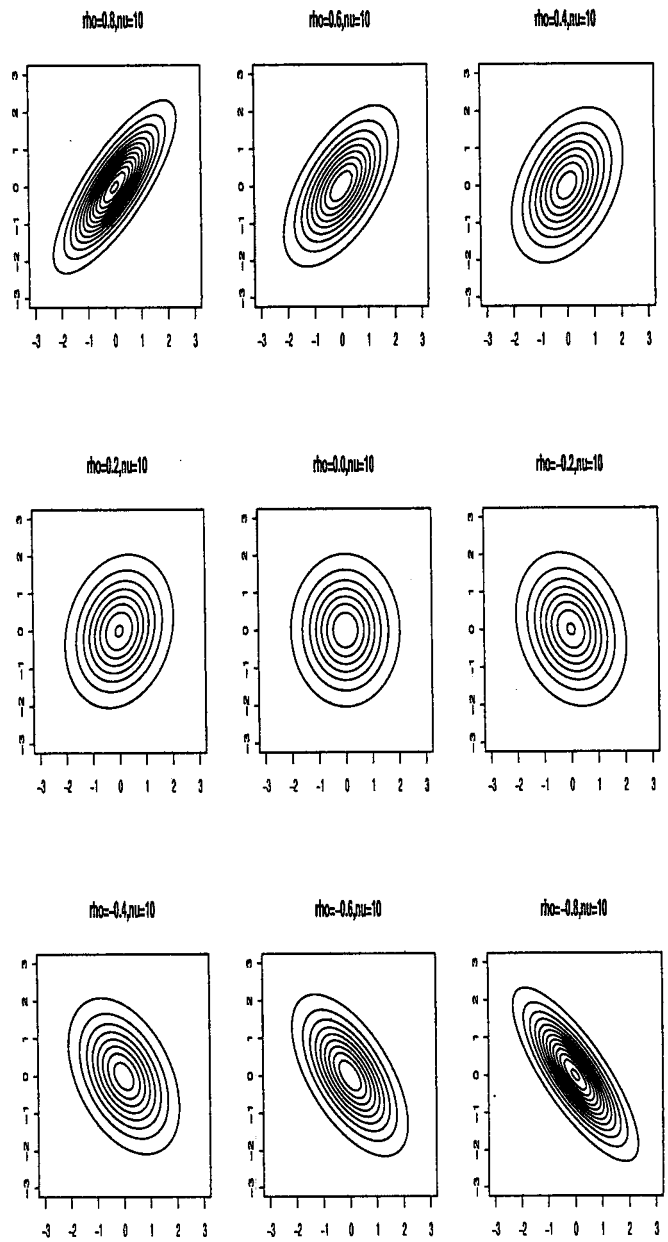

Fig. 1.3. Joint contours of (1.1) with degrees of freedom $\nu = 10$, zero means, and correlation coefficient $\rho = 0.8, 0.6, \ldots, -0.6, -0.8$

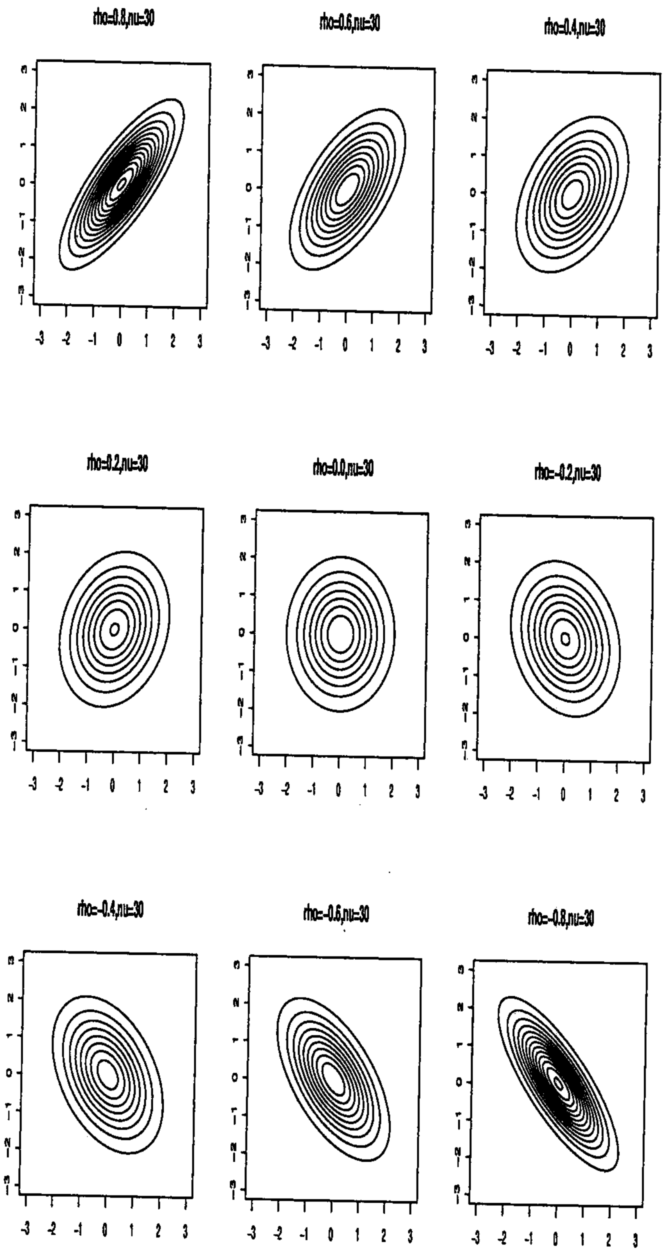

Fig. 1.4. Joint contours of (1.1) with degrees of freedom $\nu = 30$, zero means, and correlation coefficient $\rho = 0.8, 0.6, \ldots, -0.6, -0.8$

- If $\mathbf{V}^{1/2}$ is the sy

$\mathbf{V}^{1,}$

where $\mathcal{W}_p(\mathbf{\Sigma}, n)$
grees of freedom
normal distribut
the p-dimension

(Ando and Kauf
normal distributi

From representatio
joint pdf (1.1) if ar

$\mathbf{X} \mid$

$\Leftrightarrow \quad (\mathbf{a}^T$

$\Leftrightarrow \quad (\mathbf{a}^T$

and this is one of
(1962). This resul
(1.4): $\mathbf{X}$ has the jo

$\mathbf{X} \mid \mathbf{V} \sim N$

$\Leftrightarrow \quad (\mathbf{a}^T \mathbf{\Sigma} \mathbf{a})^{-1}$

$\Leftrightarrow \quad (\mathbf{a}^T \mathbf{\Sigma} \mathbf{a})^{-1}$

as noted by Lin (1
Lin (1972) obtai
tation (1.2). Let ν
dependent continu
$E(X_k \mid S^2 = s^2)$
$k = 1, \ldots, p$. Then

- $(X_1, X_2, \ldots, X_p)$
ance matrix $\mathbf{D}$,

- If $\mathbf{V}^{1/2}$ is the symmetric square root of $\mathbf{V}$, that is,

$$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V} \sim \mathcal{W}_p\left(\mathbf{R}^{-1}, \nu + p - 1\right), \tag{1.3}$$

  where $\mathcal{W}_p(\mathbf{\Sigma}, n)$ denotes the p-variate Wishart distribution with degrees of freedom n and covariance matrix $\mathbf{\Sigma}$, and if $\mathbf{Y}$ has the p-variate normal distribution with zero means and covariance matrix $\nu\mathbf{I}_p$ ($\mathbf{I}_p$ is the p-dimensional identity matrix), independent of $\mathbf{V}$, then

$$\mathbf{X} = \left(\mathbf{V}^{1/2}\right)^{-1}\mathbf{Y} + \boldsymbol{\mu} \tag{1.4}$$

  (Ando and Kaufman, 1965). This implies that $\mathbf{X} \mid \mathbf{V}$ has the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\nu\mathbf{V}^{-1}$.

1.3 Characterizations

From representation (1.2) it easily follows for any $\mathbf{a} \neq \mathbf{0}$ that $\mathbf{X}$ has the joint pdf (1.1) if and only if

$$\begin{aligned}
& \mathbf{X} \mid S^2 = s^2 \sim N\left(\boldsymbol{\mu}, s^{-2}\mathbf{\Sigma}\right) \\
\Leftrightarrow\ & \left(\mathbf{a}^T\mathbf{\Sigma}\mathbf{a}\right)^{-1/2}\mathbf{a}^T\left(\mathbf{X} - \boldsymbol{\mu}\right) \mid S^2 = s^2 \sim N\left(0, s^{-2}\right) \\
\Leftrightarrow\ & \left(\mathbf{a}^T\mathbf{\Sigma}\mathbf{a}\right)^{-1/2}\mathbf{a}^T\left(\mathbf{X} - \boldsymbol{\mu}\right) \sim t_\nu,
\end{aligned}$$

and this is one of the earliest characterization results given in Cornish (1962). This result can also be obtained by using the representation (1.4): $\mathbf{X}$ has the joint pdf (1.1) if and only if

$$\begin{aligned}
& \mathbf{X} \mid \mathbf{V} \sim N\left(\boldsymbol{\mu}, \nu V^{-1}\right) \\
\Leftrightarrow\ & \left(\mathbf{a}^T\mathbf{\Sigma}\mathbf{a}\right)^{-1/2}\mathbf{a}^T\left(\mathbf{X} - \boldsymbol{\mu}\right) \mid \mathbf{V} \sim N\left(0, \nu\left(\mathbf{a}^T\mathbf{V}^{-1}\mathbf{a}\right) \Big/ \left(\mathbf{a}^T\mathbf{\Sigma}\mathbf{a}\right)\right) \\
\Leftrightarrow\ & \left(\mathbf{a}^T\mathbf{\Sigma}\mathbf{a}\right)^{-1/2}\mathbf{a}^T\left(\mathbf{X} - \boldsymbol{\mu}\right) \sim t_\nu,
\end{aligned}$$

as noted by Lin (1972).

Lin (1972) obtained two further characterizations using the representation (1.2). Let $\nu S^2 \sim \chi^2_\nu$ and let $X_1, X_2, \ldots, X_p$ be conditionally independent continuous random variables symmetrically distributed with $E(X_k \mid S^2 = s^2) = \mu_k$ and $Var(X_k \mid S^2 = s^2) = \sigma_k^2/s^2 < \infty$ for $k = 1, \ldots, p$. Then the following characterizations are valid

- $(X_1, X_2, \ldots, X_p)^T$ has the joint pdf (1.1) with mean vector $\boldsymbol{\mu}$, covariance matrix $\mathbf{D}$, and degrees of freedom ν if and only if

$$\sum_{k=1}^{p} \frac{\left(X_k - \mu_k\right)^2}{p\sigma_k^2} \sim F_{p,\nu},$$

where $\mathbf{D}$ is a $p \times p$ diagonal matrix with its kth diagonal element equal to σ_k^2.

- In the special case $\sigma_k^2 = \sigma^2$ for all k and the conditional pdf of $X_k \mid S^2 = s^2$ is positive and differentiable for all $x \in \Re$, $(X_1, X_2, \ldots, X_p)^T$ has the joint pdf (1.1) with zero means, covariance matrix $\sigma^2 \mathbf{I}_p$, and degrees of freedom ν if and only if the joint pdf of $X_1, X_2, \ldots, X_p$ is a function of $x_1^2 + x_2^2 + \cdots + x_p^2$ only.

1.4 A Closure Property

Consider Studentizing transformations $T : \Re^n \to \Re^k$, depending on matrices $\mathbf{A}(n \times k)$, $\mathbf{B}(n \times \nu)$ and $\mathbf{\Omega}(n \times n)$, given by

$$T(\mathbf{X}) = \frac{\sqrt{\nu}\mathbf{A}^T\mathbf{X}}{\| \mathbf{B}^T\mathbf{X} \|} \tag{1.5}$$

such that $\mathbf{A}^T \mathbf{\Omega} \mathbf{B} = \mathbf{0}$. Jensen (1994) established that the class of multivariate t distributions is closed under the transform $T(\cdot)$. Specifically, assume $\mathbf{A}^T\mathbf{A} = \mathbf{I}_k$, $\mathbf{B}^T\mathbf{B} = \mathbf{I}_\nu$, and $\mathbf{X}$ is distributed according to (1.1) with zero means, correlation matrix $\mathbf{I}_n$, and degrees of freedom m. Under these assumptions, Jensen showed that $T(\mathbf{X})$ is also distributed according to (1.1) with zero means, correlation matrix $\mathbf{I}_k$, and degrees of freedom ν.

Jensen (1994) also studied the concentration properties of (1.1) via peakedness by varying its parameters. If $\mathbf{X}$ is multivariate normal, then the transformation $\mathbf{X} \to T(\mathbf{X})$ diminishes the peakedness. If, on the other hand, $\mathbf{X}$ is distributed according to (1.1) with mean vector $\mu \mathbf{1}_n$, covariance matrix $\sigma^2 \mathbf{I}_n$, and degrees of freedom m, then the transformation is peakedness-enhancing for all $m < \nu$. If $m > \nu > 2$, then the transformation serves to increase variances. For any $m > \nu > 0$ the marginal distributions are less peaked after $T(\mathbf{X})$ than before in the sense of Birnbaum (1948). If $m = \nu$, then the marginals are identical before and after $T(\mathbf{X})$, thus exhibiting identical tail behavior. If $\nu > m$ then marginals are more peaked (in the sense of Birnbaum, 1948) after applying $T(\mathbf{X})$ than before; and if $\nu > m > 2$, then $T(\mathbf{X})$ serves as a variance-diminishing transformation.

1.5 A Consistency Property

A random vector $\mathbf{X} = (X_1, \ldots, X_p)^T$ is said to have the spherical distribution if its joint pdf can be written in the form

$$g\left(\sum_{i=1}^{p} x_i^2 \middle| p\right),$$

where $g(\cdot)$ is referred to as the density generator. The p-variate t pdf (1.1) with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$ is spherical because in this case,

$$g(u) = \frac{\Gamma\left((\nu+p)/2\right)}{(\pi\nu)^{p/2}\Gamma\left(\nu/2\right)}\left(1+\frac{u}{\nu}\right)^{-(p+\nu)/2}.$$

Other examples of spherical distributions include the multivariate normal and the multivariate power exponential. A spherical distribution is said to possess the consistency property if

$$\int_{-\infty}^{\infty} g\left(\sum_{i=1}^{p+1} x_i^2 \middle| p\right) dx_{p+1} = g\left(\sum_{i=1}^{p} x_i^2 \middle| p\right) \tag{1.6}$$

for any integer p and almost all $\mathbf{x} \in \Re^p$. This consistency property ensures that any marginal distribution of $\mathbf{X}$ also belongs to the same spherical family. Kano (1994) provided several necessary and sufficient conditions for a spherical distribution to satisfy (1.6). One of the them is that g must be a mixture of normal distributions; specifically, there exists a random variable $Z > 0$, unrelated to p, such that, for any p,

$$f(u \mid p) = \int \left(\frac{z}{2\pi}\right)^{p/2} \exp\left(-\frac{uz}{2}\right) F(dz),$$

where $F(\cdot)$ denotes the cumulative distribution function (cdf) of Z. Since the multivariate t is a mixture of normal distributions (see (1.2)), it follows that it must have the consistency property. Other distributions that have the consistency property include the multivariate normal and the multivariate Cauchy. Distributions that do not share this property include the multivariate logistic, multivariate Pearson type II, multivariate Pearson type VII , and the multivariate Bessel.

1.6 Density Expansions

Fisher (1925) and later Dickey (1967a) provided expansions of the pdf

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma\left(\nu/2\right)}\left\{1+\frac{x^2}{\nu}\right\}^{-(\nu+1)/2}$$

of the univariate Student's t distribution. The expansion in the latter paper involves Appell's polynomials, and hence recurrence schemes are available for its coefficients. Specifically,

$$f(x) \quad = \quad \sqrt{\frac{1+\nu}{2\pi\nu}} \exp\left(-\frac{1+\nu}{2\nu}x^2\right) \sum_{k=0}^{\infty} Q_k \left(-\frac{1+\nu}{2\nu}x^2\right) (1+\nu)^{-k}, \tag{1.7}$$

where

$$Q_k(t) \quad = \quad P_k(t) - \frac{1}{\sqrt{\pi}} \sum_{l=0}^{k-1} Q_l(t) P_{k-l}(\Gamma). \tag{1.8}$$

Here, $P_k(t)$ are polynomials (in powers of t) satisfying

$$\sum_k P_k(t)(1+\nu)^{-k} \quad = \quad \left(1 - \frac{2t}{1+\nu}\right)^{-(1+\nu)/2} \exp(-t)$$

and $P_k(\Gamma)$ denotes the polynomial $P_k(t)$ with the powers t^r replaced by $\Gamma(r+1/2)$. Dickey (1967a) also provided an analog of (1.7) for the multivariate t pdf (1.1). It takes the same form as (1.7) with x^2 replaced by $(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{R}^{-1}(\mathbf{x}-\boldsymbol{\mu})$, $\nu+1$ replaced by $\nu+p$, and with (1.8) replaced by

$$Q_k(t) \quad = \quad P_k(t) - \frac{1}{\Gamma(p/2)} \sum_{l=0}^{k-1} Q_l(t) P_{k-l}(\Gamma_p),$$

where Γ_p indicates the substitution of $\Gamma(r+p/2)$ for t^r.

1.7 Moments

Since $\mathbf{Y}$ and S in (1.2) are independent, the conditional distribution of (X_i, X_j), given $S=s$, is bivariate normal with means (μ_i, μ_j), common variance σ^2/s^2, and correlation coefficient r_{ij}. Thus,

$$\begin{aligned} E(X_i) &= E\left[E\left(X_i \mid S=s\right)\right] \\ &= E(\mu_i) \\ &= \mu_i. \end{aligned}$$

To find the second moments, consider the classical identity

$$\begin{aligned} Cov(X_i, X_j) \quad = \quad & E\left[Cov(X_i, X_j) \mid S=s\right] \\ & + Cov\left[E\left(X_i \mid S=s\right) E\left(X_j \mid S=s\right)\right] \end{aligned}$$

for all $i, j = 1, \ldots, p$. Clearly, one has

$$E\left[Cov\left(X_i, X_j\right) \mid S = s\right] \;=\; \sigma^2 r_{ij} E\left(\frac{1}{S^2}\right)$$

and

$$Cov\left[E\left(X_i \mid S = s\right) E\left(X_j \mid S = s\right)\right] \;=\; 0.$$

If $\nu > 2$, then $E(1/S^2)$ exists and is equal to $\nu/\{\sigma^2(\nu-2)\}$. Thus, by choosing $i = j$ and $i < j$, respectively, one obtains

$$Var\left(X_i\right) \;=\; \frac{\nu}{\nu - 2}$$

and

$$Cov\left(X_i, X_j\right) \;=\; \frac{\nu}{\nu - 2} r_{ij}.$$

Hence the matrix $\mathbf{R}$ is indeed the correlation matrix as stated in definition (1.1).

In the case where $\boldsymbol{\mu} = \mathbf{0}$, the product moments of $\mathbf{X}$ are easily found by exploiting the independence of $\mathbf{Y}$ and S in (1.2). One obtains

$$\begin{aligned}
\mu_{r_1, r_2, \ldots, r_p} &= E\left[\prod_{j=1}^{p} X_j^{r_j}\right] \\
&= E\left[S^{-r}\left(\prod_{j=1}^{p} Y_j^{r_j}\right)\right] \\
&= \sigma^{-r}\nu^{r/2} E\left[\prod_{j=1}^{p} Y_j^{r_j}\right] E\left[\chi_\nu^{-r}\right],
\end{aligned}$$

provided that $r = r_1 + r_2 + \cdots + r_p < \nu/2$. In the special case where $Y_1, \ldots, Y_p$ are mutually independent, one obtains

$$\mu_{r_1, r_2, \ldots, r_p} \;=\; \sigma^{-r}\nu^{r/2} E\left[\chi_\nu^{-r}\right] \prod_{j=1}^{p} E\left[Y_j^{r_j}\right].$$

If anyone of the r_j's is odd, then the moment is zero. If all of them are even, then

$$\mu_{r_1, r_2, \ldots, r_p} \;=\; \frac{\nu^{r/2} \prod_{j=1}^{p} \{1 \cdot 3 \cdot 5 \cdots (2r_j - 1)\}}{(\nu-2)(\nu-4)\cdots(\nu-r)}, \qquad \nu > r.$$

In particular,

$$\mu_{2,0,\ldots,0} = \frac{\nu}{\nu-2}, \qquad \nu > 2,$$

$$\mu_{4,0,\ldots,0} = \frac{3\nu^2}{(\nu-2)(\nu-4)}, \qquad \nu > 4,$$

$$\mu_{2,2,0,\ldots,0} = \frac{\nu^2}{(\nu-2)(\nu-4)}, \qquad \nu > 4,$$

and

$$\mu_{2,2,2,0,\ldots,0} = \frac{\nu^3}{(\nu-2)(\nu-4)(\nu-6)}, \qquad \nu > 6.$$

1.8 Maximums

Of special interest are the moments of $Z = \max(X_1, \ldots, X_p)$ when $\mathbf{X}^T = (X_1, \ldots, X_p)$ has the t pdf (1.1) with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. These moments have applications in decision theory, particularly in the selection and estimation of the maximum of a set of parameters. It also has applications in forecasting. The problem of finding the moments of Z has been considered by Raiffa and Schlaifer (1961), Afonja (1972), and Cain (1996).

Raiffa and Schlaifer (1961) provided an expression for $E(Z-\theta)$ for the case where $p = 3$ and $\boldsymbol{\mu} = \theta \mathbf{1}_p$ (where $\mathbf{1}_p$ denotes a vector of 1's). Afonja (1972) generalized this for the general case of unequal means, variances, and correlations. We mention later a particular case of this result for $\boldsymbol{\mu} = \theta \mathbf{1}_p$. Let $\phi_p(\mathbf{y}; \mathbf{R})$ denote a p-dimensional normal pdf with zero means, unit variances, and correlation matrix $\mathbf{R}$. Also let $\mathbf{R}_i$ denote a $p \times p$ matrix with its (j, j')th element equal to $r_{i,jj'}$, where $r_{i,jj'} (j, j' \neq i)$ is the correlation between $(X_i, -X_j)$ and $(X_i, -X_{j'})$ and $r_{i,ij} = \operatorname{corr}(X_i, X_i - X_j)$. Then the kth moment of Z is given by

$$E\left(Z^k\right) = \frac{1}{\Gamma(\nu/2)} \sum_{i=1}^{p} \sum_{j=0}^{k} \binom{k}{j} \theta^{k-j} \left(\frac{\nu^2 \sigma_{ii}}{2(\nu-2)}\right)^{j/2} \Gamma\left(\frac{\nu-j}{2}\right) \mu_j\left(y_i\right), \tag{1.9}$$

where

$$\mu_j\left(y_i\right) = \int_0^{\infty} \int_0^{\infty} \cdots \int_{-\infty}^{\infty} \cdots \int_0^{\infty} \int_0^{\infty} y_i^j \phi_p\left(\mathbf{y}; \mathbf{R}_i\right) dy_p dy_{p-1} \cdots dy_i \cdots dy_2 dy_1 \tag{1.10}$$

is the marginal moment (up to a constant) of truncated normal variates. The mean and variance can be derived easily from this formula. For example,

$$E(Z) \quad = \quad \theta + \{E(W) - \theta\}\, \Gamma\left(\frac{\nu-1}{2}\right) \Big/ \Gamma\left(\frac{\nu}{2}\right),$$

where $W = \max(Y_1, \ldots, Y_p)$ for a p-variate normal random vector $\mathbf{Y}^T = (Y_1, \ldots, Y_p)$ with means equal to θ and covariance matrix $(\nu/(\nu-2))\mathbf{\Sigma}$. Afonja (1972) showed further that

$$E(W) \quad = \quad \theta + \sqrt{\frac{\nu}{\nu-2}} \sum_{i=1}^{p} \sqrt{\sigma_{ii}}\, \mu_1\left(y_i\right),$$

where $\mu_1(y_i)$ is given by (1.10) for $j = 1$.

More recently, Cain (1996) considered two forecasts F_1 and F_2 of a future variable Y where the forecast errors $X_1 = F_1 - Y$ and $X_2 = F_2 - Y$ are assumed to have the bivariate t distribution with means (μ_1, μ_2), variances (σ_1^2, σ_2^2), correlation coefficient ρ, and degrees of freedom $\nu > 2$. Cain was interested in the maximum $Z = \max(X_1, X_2)$ of the two forecast errors and whether this nonlinear function could be useful as a component of a linear combination forecast. It was shown that the pdf of Z can be written as the sum

$$f(z) \quad = \quad f_1(z) + f_2(z),$$

where

$$\begin{aligned} f_j(z) \quad &= \quad \frac{1}{\sigma_j}\sqrt{\frac{\nu}{\nu-2}}\, t_\nu\left(\sqrt{\frac{\nu}{\nu-2}}\,\frac{z-\mu_i}{\sigma_i}\right) \\ &\quad \times T_{1+\nu}\left(\frac{1+\nu\left[\frac{z-\mu_k}{\sigma_k} - \rho\frac{z-\mu_i}{\sigma_i}\right]}{\sqrt{1-\rho^2}\sqrt{\nu-2+\left(\frac{z-\mu_i}{\sigma_i}\right)^2}}\right) \end{aligned}$$

for $k = 3-j$, $j = 1, 2$. Here, t_ν and T_ν are, respectively, the pdf and the cdf of the Student's t distribution with degrees of freedom ν. Integration by parts yields that

$$E(Z) \quad = \quad \mu_1 \int_{-\infty}^{\infty} f_1(z)dz + \mu_2 \int_{-\infty}^{\infty} f_2(z)dz + \tau t_{\nu-2}\left(\frac{\mu_1-\mu_2}{\tau}\right),$$

$$Var(Z) \quad = \quad \sigma_1^2 \int_{-\infty}^{\infty} f_1(z)dz + \sigma_2^2 \int_{-\infty}^{\infty} f_2(z)dz$$

$$
\begin{aligned}
&+(\mu_1-\mu_2)^2\int_{-\infty}^{\infty} f_1(z)dz \int_{-\infty}^{\infty} f_2(z)dz \\
&+\tau(\mu_1-\mu_2)\,t_{\nu-2}\left(\frac{\mu_1-\mu_2}{\tau}\right)\int_{-\infty}^{\infty} f_2(z)dz \\
&-\tau(\mu_1-\mu_2)\,t_{\nu-2}\left(\frac{\mu_1-\mu_2}{\tau}\right)\int_{-\infty}^{\infty} f_1(z)dz \\
&+\frac{(\mu_1-\mu_2)\left(\sigma_2^2-\sigma_1^2\right)}{\tau(\nu-2)}t_{\nu-2}\left(\frac{\mu_1-\mu_2}{\tau}\right) \\
&-\tau^2 t_{\nu-2}^2\left(\frac{\mu_1-\mu_2}{\tau}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
Cov\,(Z, X_1) &= \sigma_1^2\int_{-\infty}^{\infty} f_1(z)dz + \rho\sigma_1\sigma_2\int_{-\infty}^{\infty} f_2(z)dz \\
&\quad + \frac{(\mu_1-\mu_2)\left(\sigma_1^2-\rho\sigma_1\sigma_2\right)}{\tau(\nu-2)}t_{\nu-2}\left(\frac{\mu_1-\mu_2}{\tau}\right),
\end{aligned}
$$

where $\tau=\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$. The two integrals in the above expressions can be evaluated as

$$
\int_{-\infty}^{\infty} f_1(z)dz = T_\nu\left(\frac{\mu_1-\mu_2}{\tau}\sqrt{\frac{\nu}{\nu-2}}\right)
$$

and

$$
\int_{-\infty}^{\infty} f_2(z)dz = 1-T_\nu\left(\frac{\mu_1-\mu_2}{\tau}\sqrt{\frac{\nu}{\nu-2}}\right).
$$

The expression for $Cov(Z, X_2)$ can be obtained by switching the subscripts 1 and 2. As $\nu\to\infty$, the above expressions can be reduced by replacing $t_\nu(\cdot)$ and $T_\nu(\cdot)$ by $\phi(\cdot)$ and $\Phi(\cdot)$, respectively. On the other extreme, as $\nu\to 2^+$, the expressions could be reduced by using the fact that

$$
\lim_{\nu\to 2+}\frac{|\,x\,|}{\nu-2}t_{\nu-2}(x) = \begin{cases} 0, & \text{if } x=0, \\ 1/2, & \text{if } x\neq 0, \end{cases}
$$

and

$$
\lim_{\nu\to 2+} T_{\nu-2}\left(\sqrt{\frac{\nu}{\nu-2}}x\right) = \begin{cases} 1, & \text{if } x>0, \\ 1/2, & \text{if } x=0, \\ 0, & \text{if } x<0. \end{cases}
$$

This suggests that the results for the maximum of bivariate t distributed errors may be materially different from those for bivariate normal errors.

Cain (1996) also investigated to see whether the maximum Z can provide information additional to that of F_1 and F_2 in forecasting Y via a linear combination of the form $F = \alpha + \beta_1 F_1 + \beta_2 F_2 + \gamma M$ with $\beta_1 + \beta_2 + \gamma = 1$. Cain showed that the mean squared error of F is minimized when $\gamma = 0$ and hence that M is linearly dominated by F_1 and F_2. Similar calculations reveal that the mean forecast $(F_1 + F_2)/2$ dominates M if and only if either $\mu_1 = \mu_2$ or $\sigma_1 = \sigma_2$. Evidently further investigations are in order (to consider, for example, the case of more than two forecasts).

1.9 Distribution of a Linear Function

If $\mathbf{X}$ has the p-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$, then, for any nonsingular scalar matrix $\mathbf{C}$ and for any $\mathbf{a}$, $\mathbf{CX} + \mathbf{a}$ has the p-variate t distribution with degrees of freedom ν, mean vector $\mathbf{C}\boldsymbol{\mu} + \mathbf{a}$, and correlation matrix $\mathbf{CRC}^T$. This result is of importance in applications and is similar to the corresponding result for the multivariate normal distribution.

1.10 Marginal Distributions

Let $\mathbf{X}$ possess the p-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$. Consider the partitions

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \tag{1.11}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \tag{1.12}$$

and

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}, \tag{1.13}$$

where $\mathbf{X}_1$ is $p_1 \times 1$ and $\mathbf{R}_{11}$ is $p_1 \times p_1$. Then $\mathbf{X}_1$ has the p_1-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}_1$, correlation matrix $\mathbf{R}_{11}$, and with the joint pdf given by

$$f(\mathbf{x}_1) = \frac{\Gamma((\nu + p_1)/2)}{(\pi\nu)^{p_1/2}\Gamma(\nu/2)\left|\mathbf{R}_{11}\right|^{1/2}} \times \left[1 + \frac{1}{\nu}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \mathbf{R}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)\right]^{-(\nu+p_1)/2}.$$

Moreover, $\mathbf{X}_2$ also has the $(p-p_1)$-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}_2$, correlation matrix $\mathbf{R}_{22}$, and with the joint pdf given by

$$
\begin{aligned}
f(\mathbf{x}_2) &= \frac{\Gamma\left((\nu+p-p_1)/2\right)}{(\pi\nu)^{p_1/2}\Gamma(\nu/2)\left|\mathbf{R}_{22}\right|^{1/2}} \\
&\quad \times \left[1+\frac{1}{\nu}\left(\mathbf{x}_2-\boldsymbol{\mu}_2\right)^T\mathbf{R}_{22}^{-1}\left(\mathbf{x}_2-\boldsymbol{\mu}_2\right)\right]^{-(\nu+p-p_1)/2}.
\end{aligned}
$$

1.11 Conditional Distributions

Several interesting properties have been obtained for conditional pdfs of the multivariate t distribution. If $\mathbf{X}$ has the central p-variate t distribution with degrees of freedom ν and correlation matrix $\mathbf{R}$, it then follows from Section 1.10 that the conditional pdf of $\mathbf{X}_2$ given $\mathbf{X}_1$ is given by

$$
\begin{aligned}
f(\mathbf{x}_2 \mid \mathbf{x}_1) &= \frac{\Gamma((\nu+p)/2)}{(\nu\pi)^{p_1/2}\Gamma((\nu+p_1)/2)}\frac{\left|\mathbf{R}_{11}\right|^{1/2}}{\left|\mathbf{R}\right|^{1/2}} \\
&\quad \times \frac{\left[1+(1/\nu)\mathbf{x}_1^T\mathbf{R}_{11}^{-1}\mathbf{x}_1\right]^{(\nu+p_1)/2}}{\left[1+(1/\nu)\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}\right]^{(\nu+p)/2}}. \qquad (1.14)
\end{aligned}
$$

Since

$$
\left|\mathbf{R}\right| = \left|\mathbf{R}_{11}\right|\left|\mathbf{R}_{22}-\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\right|
$$

and

$$
\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} = \mathbf{x}_1^T\mathbf{R}_{11}^{-1}\mathbf{x}_1+\mathbf{x}_{2\cdot 1}^T\mathbf{R}_{22\cdot 1}^{-1}\mathbf{x}_{2\cdot 1},
$$

where

$$
\mathbf{x}_{2\cdot 1} = \mathbf{x}_2-\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{x}_1
$$

and

$$
\mathbf{R}_{22\cdot 1} = \mathbf{R}_{22}-\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12},
$$

one can rewrite (1.14) as

$$
\begin{aligned}
f(\mathbf{x}_2 \mid \mathbf{x}_1) &= \frac{\Gamma\left((\nu+p)/2\right)}{\left\{(\nu+p_1)\pi\right\}^{(p-p_1)/2}\Gamma\left((\nu+p_1)/2\right)\left|\mathbf{R}_{22\cdot 1}\right|^{1/2}} \\
&\quad \times \left[1+\frac{1}{\nu+p_1}\frac{((\nu+p_1)/\nu)\mathbf{x}_{2\cdot 1}^T\mathbf{R}_{22\cdot 1}^{-1}\mathbf{x}_{2\cdot 1}}{1+(1/\nu)\mathbf{x}_1^T\mathbf{R}_{11}^{-1}\mathbf{x}_1}\right]^{-(\nu+p)/2} \\
&\quad \times \left[\frac{(\nu+p_1)/\nu}{1+(1/\nu)\mathbf{x}_1^T\mathbf{R}_{11}^{-1}\mathbf{x}_1}\right]^{(p-p_1)/2}. \qquad (1.15)
\end{aligned}
$$

Landenna and Ferrari (1988) noted that this conditional pdf is not a $(p-p_1)$-variate t unless the values of $\mathbf{x}_1$ are ± 1. For example, consider the special case of (1.15) for $\mathbf{R} = \mathbf{I}_p$. In this case, (1.15) becomes

$$
\begin{aligned}
f(\mathbf{x}_2 \mid \mathbf{x}_1) &= \frac{\Gamma((\nu+p)/2)}{\pi^{(p-p_1)/2}\Gamma((\nu+p_1)/2)\left(\nu + \sum_{j=1}^{p_1} x_j^2\right)^{(p-p_1)/2}} \\
&\quad \times \left[1 + \frac{1}{\nu + \sum_{j=1}^{p_1} x_j^2} \sum_{j=p_1+1}^{p} x_j^2\right]^{-(\nu+p)/2}. \qquad (1.16)
\end{aligned}
$$

When $x_j = \pm 1$, $j = 1, 2, \ldots, p_1$, (1.16) reduces to

$$
\begin{aligned}
f(\mathbf{x}_2 \mid \mathbf{x}_1) &= \frac{\Gamma((\nu+p)/2)}{\pi^{(p-p_1)/2}\Gamma((\nu+p_1)/2)(\nu+p_1)^{(p-p_1)/2}} \\
&\quad \times \left[1 + \frac{1}{\nu+p_1} \sum_{j=p_1+1}^{p} x_j^2\right]^{-(\nu+p)/2},
\end{aligned}
$$

which is the joint pdf of a central $(p-p_1)$-variate t distribution with degrees of freedom $(\nu+p_1)$ and correlation matrix $\mathbf{I}_{p-p_1}$. Landenna and Ferrari (1988) also described the manner in which the probabilities of the conditional pdf (1.15) can be expressed in terms of the probabilities of $\mathbf{x}_2$ conditioned on $\mathbf{x}_1$ taking the values ± 1.

The form of the conditional pdf (1.15) also suggests that

$$
\mathbf{Y}_1 = \mathbf{X}_1 \qquad (1.17)
$$

and

$$
\mathbf{Y}_2 = \sqrt{\frac{\nu+p_1}{\nu}} \left(1 + \frac{1}{\nu}\mathbf{X}_1^T \mathbf{R}_{11}^{-1} \mathbf{X}_1\right)^{-1/2} \left(\mathbf{X}_2 - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{X}_1\right) \qquad (1.18)
$$

are independent, that $\mathbf{Y}_1$ has the central p_1-variate t distribution with degrees of freedom ν and correlation matrix $\mathbf{R}_{11}$, and that $\mathbf{Y}_2$ has the central $(p-p_1)$-variate t distribution with degrees of freedom $\nu+p_1$ and correlation matrix $\mathbf{R}_{22\cdot 1}$. From this observation, it follows easily that the conditional expectation of $\mathbf{X}_2$ given $\mathbf{X}_1$ is linear and that $E(\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1) = \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{x}_1$. In particular,

$$
E(X_p \mid X_1 = x_1, \ldots, X_{p-1} = x_{p-1}) = \frac{1}{r_{pp}^*} \sum_{j=0}^{p-1} r_{jp}^* x_j
$$

and

$$
\begin{aligned}
& Var\,(X_p\,|X_1 = x_1, \ldots, X_{p-1} = x_{p-1}\,) \\
&= \frac{1}{r^*_{pp}} \frac{\nu}{\nu + p - 3} \left[1 + \frac{1}{\nu} \sum_{j,k=0}^{p-1} \left\{ r^*_{jk} - \frac{r^*_{jp} r^*_{kp}}{r^*_{pp}} \right\} x_j x_k \right], \quad (1.19)
\end{aligned}
$$

where r^*_{jk} is the (j,k)th element of $\mathbf{R}^{-1}$ (Bennett, 1961). It is illuminating to compare the conditional variance (1.19) with the value $1/r^*_{pp}$ corresponding to the conditional variance of the multivariate normal distribution.

Siotani (1976) generalized the result of (1.17)–(1.18) by splitting $\mathbf{X}$ into more than two sets of variates. Let

$$
\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_k \end{pmatrix} \quad (1.20)
$$

and

$$
\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1k} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{k1} & \mathbf{R}_{k2} & \cdots & \mathbf{R}_{kk} \end{pmatrix},
$$

where $\mathbf{X}_l$ is $p_l \times 1$ for $l = 1, 2, \ldots, k$ and $\mathbf{R}_{lm}$ is $p_l \times p_m$ for $l = 1, 2, \ldots, k$, $m = 1, 2, \ldots, k$. Clearly $p_1 + p_2 + \cdots + p_k = p$. Introducing the notations

$$
q_l = p_1 + p_2 + \cdots + p_l, \quad (1.21)
$$

$$
\mathbf{X}_{(l)} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_l \end{pmatrix}, \quad (1.22)
$$

$$
\mathbf{R}_{(l)} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1l} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{l1} & \mathbf{R}_{l2} & \cdots & \mathbf{R}_{ll} \end{pmatrix}, \quad (1.23)
$$

$$\mathbf{R}_{(l)}^{(l+1)} = \begin{pmatrix} \mathbf{R}_{1,l+1} \\ \mathbf{R}_{2,l+1} \\ \vdots \\ \mathbf{R}_{l,l+1} \end{pmatrix}, \tag{1.24}$$

and

$$\mathbf{R}_{l+1,l+1\cdot(l)} = \mathbf{R}_{l+1,l+1} - \mathbf{R}_{(l)}^{(l+1)^T} \mathbf{R}_{(l)}^{-1} \mathbf{R}_{(l)}^{(l+1)}, \tag{1.25}$$

Siotani showed that

$$\mathbf{Y}_1 = \mathbf{X}_1$$

and

$$\mathbf{Y}_{l+1} = \sqrt{\frac{\nu + q_l}{\nu}} \left(1 + \frac{1}{\nu} \mathbf{X}_{(l)}^T \mathbf{R}_l^{-1} \mathbf{X}_{(l)}\right)^{-1/2} \times \left(\mathbf{X}_{(l+1)} - \mathbf{R}_{(l)}^{(l+1)^T} \mathbf{R}_{(l)}^{-1} \mathbf{X}_{(l)}\right)$$

for $l = 1, \ldots, k-1$ are independent, that $\mathbf{Y}_1$ has the central p_1-variate t distribution with degrees of freedom ν and correlation matrix $\mathbf{R}_{11}$, and that $\mathbf{Y}_{l+1}$ has the central p_{l+1}-variate t distribution with degrees of freedom $(\nu + q_l)$ and correlation matrix $\mathbf{R}_{l+1,l+1\cdot(l)}$ for $l = 1, \ldots, k-1$. In the special case for $\mathbf{R} = \mathbf{I}_p$, the $\mathbf{Y}$'s can be written as

$$\mathbf{Y}_1 = \mathbf{X}_1$$

and

$$\mathbf{Y}_{l+1} = \sqrt{\frac{\nu + q_l}{\nu}} \left(1 + \frac{1}{\nu} \sum_{m=1}^{l} \mathbf{X}_m^T \mathbf{X}_m\right)^{-1/2} \mathbf{X}_{l+1}.$$

1.12 Quadratic Forms

If $\mathbf{X}$ has the p-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$, then $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}/p$ has the noncentral F distribution with degrees of freedom p and ν and noncentrality parameter $\boldsymbol{\mu}^T\mathbf{R}^{-1}\boldsymbol{\mu}/p$. See Hsu (1990) for a particular case of this result. When $\boldsymbol{\mu} = \mathbf{0}$, the distribution is central F and so $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}/(p + \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})$ has the $Beta(p/2, \nu/2)$ distribution. There are a number of problems related to quadratic forms of multivariate t that are worthy of further investigation.

1.13 F Matrix

Consider two independent random samples $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{n_1}^{(1)}$ and $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{n_2}^{(2)}$ from two different elliptical distributions (which contain multivariate t as a particular case – as already mentioned in Section 1.1). Let

$$\mathbf{S}_i \quad = \quad \sum_{k=1}^{n_i} \mathbf{x}_k^{(i)} \mathbf{x}_k^{(i)T}$$

for $i = 1, 2$. Then $\mathbf{F} = (\mathbf{S}_1/n_1)/(\mathbf{S}_2/n_2)$ is the multivariate F matrix. Hayakawa (1989) studied the asymptotic behavior of the determinant, latent roots, latent vectors, and the trace of the F matrix for an elliptical population. These results are useful in the study of the robustness of the statistics derived for testing several hypotheses about parameters of a normal population with the elliptical distribution introduced as the alternative population. Hayakawa (1989) illustrated the usefulness of the results through a multivariate t-population.

1.14 Association

The well known definition states that the random variables $X_1, \ldots, X_p$ are said to be associated if

$$Cov(f(X_1, \ldots, X_p), g(X_1, \ldots, X_p)) \quad \geq \quad 0$$

for all nondecreasing functions f, g (Esary et al., 1967). Association implies positive quadrant dependence, that is, that $\Pr\{\cap(X_i \leq x_i)\} \geq \prod_{i=1}^{p} \Pr(X_i \leq x_i)$ for all real numbers $x_1, \ldots, x_p$ (Lehmann, 1966). Jogdeo (1977) and Abdel-Hameed and Sampson (1978) established that the components of a multivariate t random vector are associated under certain conditions on correlations. More generally, the following result holds. Let $\mathbf{Z}$ be a p-variate vector with independent and real components, each having a symmetric unimodal distribution. Suppose $\mathbf{Y} = \mathbf{Z} + \mathbf{U}$, where $\mathbf{U}$ is independent of $\mathbf{Z}$ and either

(i) $\mathbf{U} = (\alpha_1 V, \ldots, \alpha_k V, \alpha_{k+1} W, \ldots, \alpha_n W)$, where (V, W) has a bivariate normal distribution centered at $\mathbf{0}$,

(ii) or $\mathbf{U} = \boldsymbol{\alpha} W$, where $\boldsymbol{\alpha}$ is an arbitrary but fixed p-variate vector and W is an arbitrary real random variable.

For $(n+1)$ independent and identically distributed (iid) copies $\mathbf{Y}_i^T = (Y_{i1}, \ldots, Y_{ip})$, $i = 0, 1, \ldots, n$ of $\mathbf{Y}$ define X_j^2, $j = 1, \ldots, p$ by

$$X_j^2 \quad = \quad \frac{nY_{0j}^2}{\sum_{i=1}^n Y_{ij}^2}.$$

Then the variables X_j^2 (or, equivalently, $\mid X_j \mid$), $j = 1, \ldots, p$ are associated.

Now, redefine $\mathbf{Y}$ as a p-variate normal random vector with zero means and covariance matrix specified by $\mathbf{\Sigma} = \{r_{ij}\sigma_i\sigma_j\}$. Let S_k^2 and S_k^{*2} be independent chi-squared random variables with degrees of freedom n and q_k, respectively, for $k = 1, \ldots, p$. Also assume that $\mathbf{X}$, S_k^2, and S_k^{*2} are mutually independent. Then, as a consequence of the above general result, one could provide the following assertions about bivariate and trivariate t vectors

- For $p = 2$, the random variables

  $$(X_1, X_2) \quad = \quad \left(\frac{\mid Y_1 \mid}{\sqrt{S_1^2 + S_1^{*2}}}, \frac{\mid Y_2 \mid}{\sqrt{S_2^2 + S_2^{*2}}} \right)$$

  are associated.
- For $p = 3$, if $\prod_{i<j} \operatorname{sign}(\lambda_{ij}) \le 0$, where $\mathbf{\Lambda} = \{\lambda_{ij}\} = \mathbf{\Sigma}^{-1}$, then the random variables

  $$(X_1, X_2, X_3) \quad = \quad \left(\frac{\mid Y_1 \mid}{\sqrt{S_1^2 + S_1^{*2}}}, \frac{\mid Y_2 \mid}{\sqrt{S_2^2 + S_2^{*2}}}, \frac{\mid Y_3 \mid}{\sqrt{S_3^2 + S_3^{*2}}} \right)$$

  are associated.

1.15 Entropy

The entropy of a continuous random vector $\mathbf{X}$ may be regarded as a descriptive quantity, just as the median, mode, variance, and the coefficient of skewness may be regarded as descriptive parameters. The entropy is a measure of the extent to which a multivariate distribution is concentrated on a few points or dispersed over many points. Thus, the entropy is a measure of dispersion, somewhat like the standard deviation in the univariate case.

Mathematically, the entropy of $\mathbf{X}$ is defined by

$$H(\mathbf{X}) \quad = \quad E\left[-\log f(\mathbf{X})\right]$$

$$= -\int f(\mathbf{x}) \log f(\mathbf{x})\, d\mathbf{x}. \tag{1.26}$$

Guerrero-Cusumano (1996a) derived the forms of this for the multivariate t distribution. For a central p-variate t, it turns out that

$$\begin{aligned} H(\mathbf{X};\mathbf{R}) &= \frac{1}{2}\log|\mathbf{R}| + \log\left[\frac{(\nu\pi)^{p/2}}{\Gamma(p/2)} B\left(\frac{p}{2},\frac{\nu}{2}\right)\right] \\ &\quad + \frac{\nu+p}{2}\left[\psi\left(\frac{\nu+p}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right], \end{aligned} \tag{1.27}$$

where $\psi(t) = d\log\Gamma(t)/dt$ denotes the digamma function. Note that (1.27) can reexpressed as $H(\mathbf{X}) = 1/2 \mid \mathbf{R} \mid + \Phi(\nu,p)$, where $\Phi(\nu,p)$ is a constant that depends only on ν and p. Table 1 in Guerrero-Cusumano (1996a) tabulates $\Phi(\nu,p)$ for $\nu = 1(1)35$ and $p = 1(1)5$. The following is an abridged version of the table.

Constant Φ for $H(\mathbf{X}) = 1/2 \mid \mathbf{R} \mid + \Phi(\nu,p)$

ν	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
1	2.53102	4.83788	7.06205	9.24381	11.3999
2	1.96028	3.83788	5.67306	7.48261	9.27502
3	1.77348	3.50454	5.20997	6.89826	8.57432
4	1.68176	3.33788	4.97687	6.60362	8.22121
5	1.62750	3.23788	4.83602	6.42500	8.00685
6	1.59172	3.17121	4.74153	6.30474	7.86226
7	1.56638	3.12359	4.67368	6.21809	7.75785
8	1.54750	3.08788	4.62257	6.15261	7.67878
9	1.53289	3.06010	4.58266	6.10135	7.61677
10	1.52126	3.03788	4.55062	6.06010	7.56678

The particular case of (1.27) for $\nu = 1$ gives the entropy for the multivariate Cauchy distribution

$$\begin{aligned} H(\mathbf{X};\mathbf{R}) &= \frac{1}{2}\log|\mathbf{R}| + \log\left[\frac{\pi^{p/2}}{\Gamma(p/2)} B\left(\frac{p}{2},\frac{1}{2}\right)\right] \\ &\quad + \frac{1+p}{2}\left[\psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right)\right]. \end{aligned}$$

As $\nu \to \infty$, (1.27) converges to the entropy of the normal distribution given by

$$H\left(\mathbf{X};\mathbf{R}\right) \;=\; \frac{p}{2}\log(2e\pi)+\frac{1}{2}\log\left|\mathbf{R}\right|. \qquad (1.28)$$

The sampling properties of (1.27) will be discussed in Chapter 9.

For the noncentral p-variate t, (1.26) takes the general form

$$H\left(\mathbf{X};\mathbf{R}\right) \;=\; \frac{1}{2}\log\left|\mathbf{R}\right|+\log\left[\frac{(\nu\pi)^{p/2}}{\Gamma\left(p/2\right)}B\left(\frac{p}{2},\frac{\nu}{2}\right)\right]+\frac{\nu+p}{2}M\left(\nu,p,\Delta\right), \qquad (1.29)$$

where $\Delta=\boldsymbol{\mu}^T\mathbf{R}^{-1}\boldsymbol{\mu}$ and $M(\nu,p,\Delta)$ is given by

$$M\left(\nu,p,\Delta\right) \;=\; \exp\left(-\frac{\Delta}{2}\right)\sum_{j=0}^{\infty}\frac{1}{j!}\left\{\psi\left(\frac{\nu+p+2j}{2}\right)-\psi\left(\frac{\nu}{2}\right)\right\}.$$

Setting $\nu=1$ in (1.29), one can obtain the entropy of the noncentral p-variate Cauchy distribution. In the case $p=1$, (1.29) coincides with the entropies for the univariate Student's t and Cauchy distributions given, for example, in Lazo and Rathie (1978).

Zografos (1999) provided a maximum entropy characterization of (1.1). The maximum entropy principle suggests to approximate the unknown pdf of $\mathbf{X}$ by the model that maximizes (1.26) subject to the constraints that define the class of pdfs considered. Jaynes (1957) asserted that the maximum entropy distribution, obtained by this constrained maximization problem, "is the only unbiased assignment we can make; to use any other would amount to an arbitrary assumption of information which by hypothesis we do not have." Zografos (1999) showed that (1.1) is the solution to maximizing $E[-\log f(\mathbf{X})]$ subject to the constraint

$$E\left[\log\left\{1+\frac{1}{\nu}\left(\mathbf{X}-\boldsymbol{\mu}\right)^T\mathbf{R}^{-1}\left(\mathbf{X}-\boldsymbol{\mu}\right)\right\}\right] \;=\; w\left(\frac{p+\nu}{2};\frac{p}{2}\right),$$

where $w(x;\alpha)=\psi(x)-\psi(x-\alpha)$, $x>\alpha$, and $\psi(\cdot)$ denotes the digamma function. For further discussion of maximum entropy methods, see Fry (2002).

1.16 Kullback-Leibler Number

The mutual information of a continuous random vector $\mathbf{X}$ with joint pdf $f(\mathbf{x})$ and marginal pdfs $f(x_i)$, $i=1,\ldots,p$ is defined by

$$T\left(\mathbf{X}\right) \;=\; E\left[-\log\left\{\frac{f\left(\mathbf{X}\right)}{f\left(x_1\right)\cdots f\left(x_p\right)}\right\}\right] \qquad (1.30)$$

with the domain of variation given by $0 \leq T(\mathbf{X}) < \infty$. (The reader should not confuse this with the transformation $T(\mathbf{X})$ given in (1.5).) The quantity (1.30) can be considered a measure of dependence (Joe, 1989). The larger the $T(\mathbf{X})$, the higher the dependence among the variables X_i, $i = 1, \ldots, p$. Naturally, $T(\mathbf{X}) = 0$ implies that the variables are independent; this latter statement follows from the fact that T is a special case of the Kullback-Leibler number, $KL(f, g)$ (Kullback, 1968). When the variables of $\mathbf{X}$ are multivariate normal with covariance matrix $\mathbf{\Sigma}$, it is easy to compute $T(\mathbf{X})$ as the difference between entropies given by (1.28); specifically,

$$T(\mathbf{X}; \mathbf{\Sigma}) = H(\mathbf{X}; \mathbf{\Sigma}) - H(\mathbf{X}; \mathbf{D}),$$

where $\mathbf{D}$ is a diagonal matrix corresponding to $\mathbf{\Sigma}$ with the elements $\sigma_{11}, \ldots, \sigma_{pp}$. This is due to the well known fact that uncorrelatedness implies independence in the normal case. This fact also implies that $T(\mathbf{X}; \mathbf{I}) = 0$. In general, for any member of an elliptical family of distributions, this is not true; in other words, uncorrelatedness does not imply that $T(\mathbf{X}) = 0$. The mutual information attempts to summarize in a single number the whole dependence structure of the multivariate distribution of $\mathbf{X}$.

Guerrero-Cusumano (1996b) derived the form of (1.30) for the multivariate t distribution. For a central p-variate t, it turns out that

$$T(\mathbf{X}) = \Omega - \frac{1}{2} \log |\mathbf{R}|, \tag{1.31}$$

where Ω is given by

$$\begin{aligned}\Omega = {} & \log \left\{ \frac{\Gamma(p/2)}{\pi^{p/2}} \frac{B^p\left(\frac{1+\nu}{2}, \frac{1}{2}\right)}{B\left(\frac{p+\nu}{2}, \frac{p}{2}\right)} \right\} + \frac{p(1+\nu)}{2} \left\{ \psi\left(\frac{1+\nu}{2}\right) - \psi\left(\frac{\nu}{2}\right) \right\} \\ & - \frac{p+\nu}{2} \left\{ \psi\left(\frac{p+\nu}{2}\right) - \psi\left(\frac{\nu}{2}\right) \right\}. \end{aligned} \tag{1.32}$$

It is easy to see that $\Omega \to 0$ as $\nu \to \infty$. The mutual information for the multivariate normal distribution with correlation matrix $\mathbf{R}$ is given by $-(1/2) \log \mid \mathbf{R} \mid$ (Kullback, 1968). The particular case of (1.31) for $\nu = 1$ gives the mutual information for the multivariate Cauchy distribution with Ω taking the simpler form

$$\Omega = \log \left\{ \frac{8^p}{\pi^{p/2}} \frac{\Gamma\left(p + \frac{1}{2}\right)}{\Gamma\left(\frac{1+p}{2}\right)} \right\} - \frac{1+p}{2} \left\{ \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) \right\}.$$

Table 1 in Guerrero-Cusumano (1996b) provides values of (1.32) for a range of ν and p. The following is an abridged version.

Constant Ω for $T(\mathbf{X}) = \Omega - (1/2)\log|\mathbf{R}|$

ν	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
1	0	0.4196180	0.949615	1.530690	2.141170
2	0	0.2927000	0.705474	1.184010	1.704100
3	0	0.2254360	0.565424	0.975130	1.431820
4	0	0.1835450	0.473177	0.832265	1.240460
5	0	0.1548760	0.407380	0.727338	1.096790
6	0	0.1339950	0.357917	0.646600	0.984235
7	0	0.1180970	0.319304	0.582368	0.893344
8	0	0.1055830	0.288289	0.529959	0.818244
9	0	0.0954730	0.262813	0.486337	0.755056
10	0	0.0871342	0.241503	0.449434	0.701101

Figures 1.5 and 1.6 graph $T(\mathbf{X})$ in (1.31) for $p = 2$ and $p = 4$, respectively. The correlation matrix $\mathbf{R}$ is taken to have the equicorrelation structure $r_{ij} = \rho$, $i \neq j$. It is interesting to see the "dale-shaped" three-dimensional plot. The figures show that, as one moves toward the center of the "dale," the dependence among the variables decreases, and, as one moves away from the center, the dependence increases.

For the normal case, Linfoot (1957) and Joe (1989) suggested a parameterization for $T(\mathbf{X})$ to make it comparable to a correlation coefficient. They defined the induced correlation coefficient based on the mutual information as

$$\rho_I = \sqrt{1 - \exp\{-2T(\mathbf{X})\}}. \tag{1.33}$$

Guerrero-Cusumano (1998) suggested a similar measure for the multivariate t distribution referred to as the *dependence coefficient.* It is given by

$$\rho_I = \sqrt{1 - |\mathbf{R}|\exp(-2\Omega)}. \tag{1.34}$$

The dependence coefficient is a quantification of dependence among the p variables of $\mathbf{X}$. This follows from the fact that independence implies

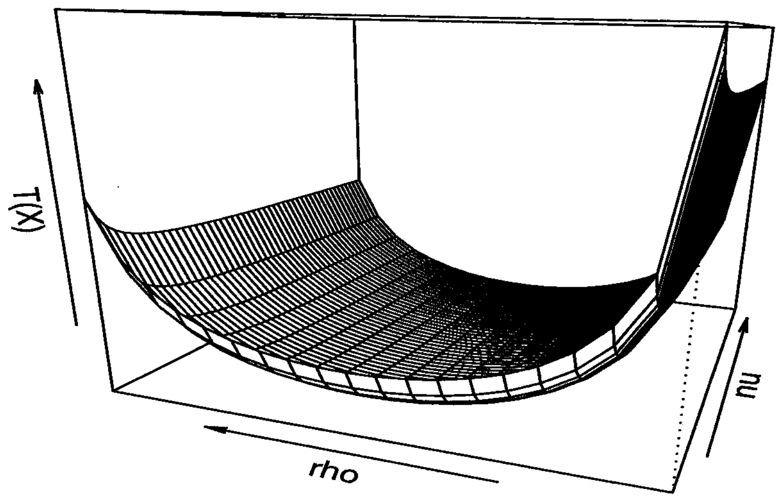

Fig. 1.5. Mutual information, (1.31), for $p = 2$

$\rho_I = 0$ and that $T(\mathbf{X}) = \infty$ implies $\rho_I = 1$. When $\nu \to \infty$, (1.34) coincides with (1.33).

The sampling properties of (1.31) will be discussed in Chapter 9.

1.17 Rényi Information

Since the concept of Rényi information is not widely available in the literature, we provide here a brief discussion of the concept. Rényi information of order λ for a continuous random variable with pdf f is defined as

$$\mathcal{I}_R(\lambda) \; := \; \frac{1}{1-\lambda} \log \left(\int f^{\lambda}(x) dx \right) \qquad (1.35)$$

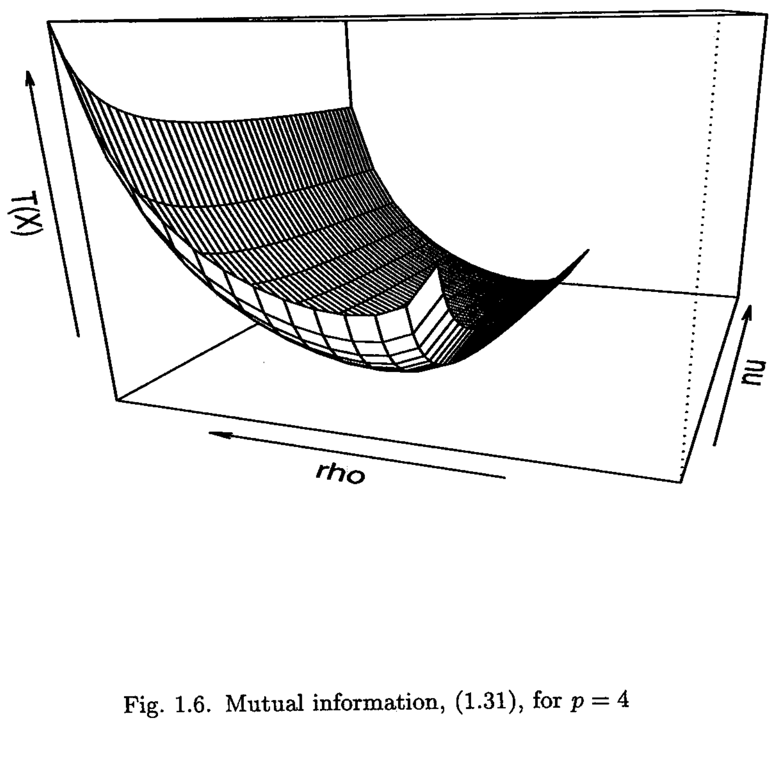

Fig. 1.6. Mutual information, (1.31), for $p = 4$

for $\lambda \neq 1$. Its value for $\lambda = 1$ is taken as the limit

$$\begin{aligned}
\mathcal{I}_R(1) &:= \lim_{\lambda \to 1} \mathcal{I}_R(\lambda) \\
&= -\int f(x) \log (f(x))\, dx \\
&= -E\left[\log f(X)\right],
\end{aligned}$$

which is the well known Shannon entropy. Rényi's (1959, 1960, 1961) generalization of the Shannon entropy allows for "different averaging of probabilities" via λ. Sometimes (1.35) is also referred to as the spectrum of Rényi information. Rényi information finds its applications as a measure of complexity in areas of physics, information theory, and engineering to describe many nonlinear dynamical or chaotic systems (Kurths et al., 1995) and in statistics as certain appropriately scaled test statistics (Rényi distances or relative Rényi information) for testing hypotheses in parametric models (Morales et al., 1997). The gradient $\mathcal{I}'_R(\lambda) = d\mathcal{I}_R(\lambda)/d\lambda$ also conveys useful information. In fact, a direct calculation based on (1.35) – assuming that the integral $\int f^\lambda(x)dx$ is well defined and differentiation operations are legitimate – shows that

$$\begin{aligned}
\mathcal{I}'_R(1) &= \lim_{\lambda\to 1}\left[(1-\lambda)\frac{\int f^\lambda(x)\log f(x)dx}{\int f^\lambda(x)dx} + \log\left(\int f^\lambda(x)dx\right)\right] \Big/ (1-\lambda)^2 \\
&= -\frac{1}{2}\lim_{\lambda\to 1}\left\{\frac{\int f^\lambda(x)\log^2 f(x)dx}{\int f^\lambda(x)dx} - \left(\frac{\int f^\lambda(x)\log f(x)dx}{\int f^\lambda(x)dx}\right)^2\right\} \\
&= -\frac{1}{2}Var\left[\log f(X)\right].
\end{aligned}$$

In other words, the gradient of Rényi information at $\lambda = 1$ is simply the negative half of the variance of the log-likelihood compared to the entropy as the negative of the expected log-likelihood. Thus, the variance of the log-likelihood $\mathcal{I}_f := 2\mathcal{I}'_R(1)$ measures the intrinsic shape of the distribution. This can be seen by observing that $\mathcal{I}_f$, where $f(x) = (1/\sigma)g((x-\mu)/\sigma)$. In fact, according to Bickel and Lehmann (1975), it can serve as a measure of the shape of a distribution. In the case where $f(x)$ has a finite fourth moment, it plays a similar role as a kurtosis measure in comparing the shapes of various frequently used densities and measuring the heaviness of tails, but it measures more than what kurtosis measures.

Rényi information of order λ for a p-variate random vector with joint pdf $\mathbf{x}$ is defined as

$$\mathcal{I}_R(\lambda) := \frac{1}{1-\lambda}\log\left(\int f^\lambda(x_1,\ldots,x_p)\,dx_1\cdots dx_p\right). \tag{1.36}$$

The gradient $\mathcal{I}'_R(\lambda)$ and the measure $\mathcal{I}_f$ are defined similarly.

Song (2001) provided a comprehensive account of $\mathcal{I}_R(\lambda)$, $\mathcal{I}'_R(\lambda)$, and $\mathcal{I}_f$ for well known univariate and multivariate distributions. For the univariate Student's t distribution with degrees of freedom ν, it can be shown for $\lambda > 1/(1+\nu)$ that

$$\mathcal{I}_R(\lambda) = \frac{1}{1-\lambda}\log\left\{\frac{B\left((\nu\lambda+\lambda-1)/2, 1/2\right)}{B^\lambda\left(\nu/2, 1/2\right)}\right\} + \frac{1}{2}\log(\nu),$$

$$
\begin{aligned}
\mathcal{I}_R'(\lambda) &= \Bigg[\log\left\{\frac{B\left((\nu\lambda+\lambda-1)/2, 1/2\right)}{B\left(\nu/2, 1/2\right)}\right\} \\
&\quad + \frac{(1-\lambda)(1+\nu)}{2}\psi\left(\frac{\nu\lambda+\lambda-1}{2}\right) \\
&\quad - \frac{(1-\lambda)(1+\nu)}{2}\psi\left(\frac{(1+\nu)\lambda}{2}\right)\Bigg]\Big/(1-\lambda)^2,
\end{aligned}
$$

and

$$
\mathcal{I}_f(\nu) = \frac{(1+\nu)^2}{4}\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{1+\nu}{2}\right)\right\}.
$$

Using tables in Abramowitz and Stegun (1965), one obtains the particular values

$$
\begin{aligned}
\mathcal{I}_f(1) &= \frac{\pi^2}{3}, \\
\mathcal{I}_f(2) &= 9 - \frac{3\pi^2}{4}, \\
\mathcal{I}_f(3) &= \frac{4\pi^2}{3} - 12, \\
\mathcal{I}_f(4) &= \frac{775}{36} - \frac{25\pi^2}{12}, \\
\mathcal{I}_f(5) &= 3\pi^2 - \frac{115}{4}.
\end{aligned}
$$

It is interesting to note that the measure $\mathcal{I}_f(\nu)$ decreases as ν increases, which makes sense since the tails become lighter as ν increases. In fact, it can be shown, using asymptotic formulas for the trigamma function, that $\lim_{\nu\to\infty}\mathcal{I}_f(\nu) = 1/2$, which corresponds to the measure $\mathcal{I}_f(\nu)$ for the normal distribution.

For the central p-variate t distribution with correlation matrix $\mathbf{R}$ and degrees of freedom ν, it can be shown for $\lambda > p/(p+\nu)$ that

$$
\begin{aligned}
\mathcal{I}_R(\lambda) &= \frac{1}{1-\lambda}\log\left\{\frac{B\left((\nu\lambda+p\lambda-p)/2, p/2\right)}{B^{\lambda}\left(\nu/2, p/2\right)}\right\} + \frac{1}{2}\log\left\{(\nu\pi)^p \mid \mathbf{R} \mid\right\} \\
&\quad - \log\Gamma\left(\frac{p}{2}\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{I}_R'(\lambda) &= \Bigg[\log\left\{\frac{B\left((\nu\lambda+p\lambda-1)/2, p/2\right)}{B\left(\nu/2, p/2\right)}\right\} \\
&\quad + \frac{(1-\lambda)(p+\nu)}{2}\psi\left(\frac{\nu\lambda+p\lambda-p}{2}\right)
\end{aligned}
$$

$$-\frac{(1-\lambda)(p+\nu)}{2}\psi\left(\frac{(p+\nu)\lambda}{2}\right)\Bigg]\Bigg/(1-\lambda)^2,$$

and

$$\mathcal{I}_f(\nu) \quad = \quad \frac{(p+\nu)^2}{4}\left\{\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{p+\nu}{2}\right)\right\}.$$

For $p = 1$, these expressions reduce to those derived for the Student's t distribution.

1.18 Identities

In one of the earliest papers on the subject, Dickey (1965, 1968) provided two multidimensional-integral identities involving the multivariate t distribution. This first identity expresses a moment of a product of multivariate t densities of the form (1.1) as an integral of dimension 1 less than the number of factors. Consider the product

$$g(\mathbf{x}) \quad = \quad \prod_{k=1}^{K}\left[1+(\mathbf{x}-\boldsymbol{\mu}_k)^T\mathbf{R}_k(\mathbf{x}-\boldsymbol{\mu}_k)\right]^{-\nu_k/2}, \qquad (1.37)$$

where each $\mathbf{R}_k \geq 0$ and $\nu_k > 0$, and so each term may not have a finite integral. The identity seeks an expression for the complete p-dimensional integral of $s \cdot g$, where $s(\mathbf{x})$ is a polynomial in the coordinates of $\mathbf{x}$. Let $\mathbf{Y}$ be a p-variate normal random vector with the covariance matrix and mean vector given by

$$\mathbf{D}_u^{-1} \quad = \quad \left(\sum_{k=1}^{K}u_k\mathbf{R}_k\right)^{-1}$$

and

$$\bar{\boldsymbol{\mu}}_u \quad = \quad \mathbf{D}_u^{-1}\sum_{k=1}^{K}u_k\mathbf{R}_k\boldsymbol{\mu}_k,$$

respectively. For given constants $c_k > 0$, $k = 1, \ldots, K$, let $u_. = \sum_{k=1}^{K} c_k u_k$ and $u_k = v_k u_.$. Then the quantity defined by $N_{s|u} = E(s(\mathbf{Y}))$ can be expanded as a polynomial in $1/u_.$ as

$$N_{s|u} \quad = \quad \sum_j h_j(v_1, \ldots, v_K)\, u_.^{-j}.$$

Given this terminology, the identity can now be expressed as

$$\begin{aligned}
&\int_{\Re^p} s(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x} \\
&= \frac{K_0}{c_K} \sum_j 2^{-j} \Gamma\left(\frac{\nu_. - p}{2} - j\right) \int_\sigma |\mathbf{D}_v|^{-1/2} h_j(v_1, \ldots, v_K) \\
&\qquad \times \left(\prod_{k=1}^{K} v_k^{\nu_k/2-1}\right) W_v^{j-(\nu_. - p)/2} dv_1 \cdots dv_{K-1}, \qquad (1.38)
\end{aligned}$$

where

$$K_0 = \pi^{p/2} \Big/ \prod_{k=1}^{K} \Gamma(\nu_k/2),$$

$$\nu_. = \sum_{k=1}^{K} \nu_k,$$

$$\mathbf{D}_v = \sum_{k=1}^{K} v_k \mathbf{R}_k,$$

$$W_v = \sum_{k=1}^{K} v_k \left\{1 + \boldsymbol{\mu}_k^T \mathbf{R}_k \boldsymbol{\mu}_k\right\} - \left(\sum_{k=1}^{K} v_k \mathbf{R}_k \boldsymbol{\mu}_k\right)^T \mathbf{D}_v^{-1} \left(\sum_{k=1}^{K} v_k \mathbf{R}_k \boldsymbol{\mu}_k\right),$$

and σ is the simplex

$$\sigma = \left\{(v_1, \ldots, v_K) : \sum_{k=1}^{K} c_k v_k = 1, \quad v_k > 0\right\}.$$

This identity has applications to inference concerning the location parameters of a multivariate normal distribution. In the particular case $K = 2$, $\mathbf{R}_k = \gamma_k \mathbf{I}_p$, and $s \equiv 1$, (1.38) reduces to

$$\begin{aligned}
\int_{\Re^p} g(\mathbf{x})\, d\mathbf{x} &= C \gamma_2^{(\nu_. - p)/2} B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \\
&\qquad \times F_1\left(\frac{\nu_1}{2}; \frac{\nu_. - p}{2}, \frac{\nu_. - p}{2}, \frac{\nu_.}{2}; z_1, z_2\right), \qquad (1.39)
\end{aligned}$$

where

$$C = \frac{\Gamma((\nu_. - p)/2)}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \frac{\pi^{p/2}}{\gamma_1^{\nu_1/2} \gamma_2^{\nu_2/2}},$$

F_1 is Appell's hypergeometric function of two variables defined by

$$\begin{aligned} & F_1\left(\alpha;\beta,\beta';\gamma;x,y\right) \\ = \; & \frac{1}{B\left(\alpha,\gamma-\alpha\right)}\int_0^1 t^{\alpha-1}(1-t)^{\gamma-\alpha-1}(1-tx)^{-\beta}(1-ty)^{-\beta'}\,dt \end{aligned} \tag{1.40}$$

(see, for example, Erdélyi et al., 1953), and z_1 and z_2 are the two real roots of the equation

$$z^2+\left(\gamma_2 \parallel \boldsymbol{\mu}_2-\boldsymbol{\mu}_1 \parallel^2+\frac{\gamma_2}{\gamma_1}-1\right)z-\gamma_2 \parallel \boldsymbol{\mu}_2-\boldsymbol{\mu}_1 \parallel^2 \;=\; 0.$$

The integral (1.39) is proportional to a multivariate generalization of the Behren-Fisher density. For an asymptotic expansion of (1.37) in powers of ν_k, see Dickey (1967a).

The second identity given by Dickey (1968) – see also Dickey (1966b) – expresses the density of a linear combination of independently distributed multivariate t vectors as an integral of dimension 1 less than the number of summands. Consider the r-variate vector $\boldsymbol{\delta}$ formed by the linear combination

$$\boldsymbol{\delta} \;=\; \sum_{k=1}^{K}\mathbf{B}_k\mathbf{X}_k,$$

where $\mathbf{X}_k$ are independent q_k-variate standard t random vectors with zero means, covariance matrix $\mathbf{I}_{q_k}$, and degrees of freedom ν_k. Dickey (1968) showed that $\boldsymbol{\delta}$ has the representation

$$\boldsymbol{\delta} \;=\; \sqrt{\sum_{k=1}^{K}\nu_k U_k^{-1}\mathbf{B}_k\mathbf{B}_k^T\mathbf{Y}},$$

where U_k are independent chi-squared random variables with degrees of freedom ν_k and $\mathbf{Y}$ is an independent r-variate standard normal vector. As a consequence, $\boldsymbol{\delta}$ has the further representation

$$\boldsymbol{\delta} \;=\; \sqrt{\sum_{k=1}^{K}\nu_k V_k^{-1}\left(\nu_k/\nu.\right)\mathbf{B}_k\mathbf{B}_k^T\mathbf{W}},$$

where $\nu. \;=\; \sum_{k=1}^{K}\nu_k$, $V_k \;=\; U_k/\sum_{j=1}^{K}U_j$ and $\mathbf{W}$ is an independent r-variate standard t vector with degrees of freedom $\nu.$. If the matrix $\sum\mathbf{B}_k\mathbf{B}_k^T$ is nonsingular, the distribution of $\boldsymbol{\delta}$ is nondegenerate with the joint pdf

$$f(\boldsymbol{\delta}) = C\int_\sigma \left(\prod_{k=1}^{K} v_k^{\nu_k/2-1}\right)\left\{1+\boldsymbol{\delta}^T\left(\sum_{k=1}^{K}(\nu_k/v_k)\,\mathbf{B}_k\mathbf{B}_k^T\right)^{-1}\boldsymbol{\delta}\right\}^{-(\nu.+r)/2} \Big/ \sqrt{\sum_{k=1}^{K}(\nu_k/v_k)\,\mathbf{B}_k\mathbf{B}_k^T}\, dv_1\cdots dv_{K-1}, \qquad (1.41)$$

where

$$C = \frac{\Gamma\left((\nu.+r)/2\right)}{\pi^{r/2}\Gamma\left(\nu_1/2\right)\cdots\Gamma\left(\nu_K/2\right)}$$

and as above

$$\sigma = \left\{(v_1,\ldots,v_K):\sum_{k=1}^{K}c_k v_k = 1,\quad v_k>0\right\}.$$

This identity has applications in Behrens-Fisher problems. The version of (1.41) for $K=2$ and $\mathbf{B}_k=\beta_k$ is

$$f(\boldsymbol{\delta}) = CB\left(\frac{\nu_1+p}{2},\frac{\nu_2+p}{2}\right) \times F_1\left(\frac{\nu_1+p}{2};\frac{\nu.+p}{2},\frac{\nu.+p}{2};\frac{\nu.}{2}+p;z_1,z_2\right),$$

where

$$C = \frac{\Gamma\left((\nu.+p)/2\right)}{\pi^{p/2}\Gamma\left(\nu_1/2\right)\cdots\Gamma\left(\nu_2/2\right)}\left(\nu_1\beta_1^2\right)^{\nu_1/2}\left(\nu_2\beta_2^2\right)^{-(\nu_1+p)/2},$$

F_1 is Appell's hypergeometric function as defined in (1.40), and z_1 and z_2 are the two real roots of the equation

$$z^2+\left(\frac{\|\boldsymbol{\delta}\|^2}{\nu_2\beta_2^2}+\frac{\nu_1\beta_1^2}{\nu_2\beta_2^2}-1\right)z-\frac{\|\boldsymbol{\delta}\|^2}{\nu_2\beta_2^2} = 0.$$

This special case is essentially equivalent to the two-factor version of (1.38). Moreover, (1.41) is a generalization of Ruben's (1960) integral representation (in the univariate case) for the usual Behrens-Fisher densities.

1.19 Some Special Cases

A number of special cases of (1.1) have been studied in the literature with great detail. Cornish (1954), in his early paper, considered the special case of (1.1) when $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{R}$ is given by the equicorrelation matrix

$$\mathbf{R} = \begin{pmatrix} 1 & -1/p & \dots & -1/p \\ -1/p & 1 & \dots & -1/p \\ \vdots & \vdots & \dots & \vdots \\ -1/p & -1/p & \dots & 1 \end{pmatrix}.$$

The following interesting properties were established

- $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}$ has the noncentral F distribution with degrees of freedom p and ν.
- $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}$ has the Fisher's z distribution with degrees of freedom $p-q$ and ν – when $\mathbf{X}$ is subject to the linearly independent homogeneous conditions represented by the equation $\mathbf{SX} = \mathbf{0}$, where $\mathbf{S}$ is of order $q \times p$ and rank $q < p$.
- The cdf of the quadratic form $Q = \mathbf{X}^T\mathbf{AX}$ when $\mathbf{A}$ is of rank $q \le p$ is given by

$$\frac{\Gamma\left((\nu+q)/2\right)}{(\pi\nu)^{q/2}\Gamma\left(\nu/2\right)} \int \cdots \int \left(1 + \frac{\mathbf{x}_1^T\mathbf{x}_1}{\nu}\right)^{-(\nu+q)/2} d\mathbf{x}_1,$$

  where $\mathbf{x}_1^T = (x_1, \dots, x_q)$ and the domain of integration is defined by

$$\sum_{i=1}^{q} \lambda_i x_i^2 \ge Q,$$

  where λ_i are the roots of the equation $\mid \lambda\mathbf{R}^{-1} - \mathbf{A} \mid = 0$ or, alternatively, the latent roots of the matrix $\mathbf{RA}$. Consequently, the distribution of $\mathbf{X}^T\mathbf{AX}$ is Fisher's z with degrees of freedom q and ν if and only if the nonzero latent roots of $\mathbf{RA}$ are all equal to unity.
- If the distribution of $\mathbf{X}$ is partitioned as in (1.11)–(1.13), then

$$E\left(\mathbf{X}_1 \mid \mathbf{X}_2\right) = -\mathbf{R}_{11}^{-1}\mathbf{R}_{22}\mathbf{x}_2,$$

  and

$$Var\left(\mathbf{X}_1 \mid \mathbf{X}_2\right) = \frac{\nu + \mathbf{x}_2^T\left(\mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\right)\mathbf{x}_2}{\nu + p - p_1 - 2}\mathbf{R}_{11}^{-1}.$$

  In the particular case $p_1 = 1$,

$$E\left(X_1 \mid \mathbf{X}_2\right) = -\frac{1}{2}\sum_{j=2}^{p} x_j,$$

$$
\begin{aligned}
Var\,(X_1 \mid \mathbf{X}_2) &= \frac{(p+1)\nu}{2p(\nu+p-3)} + \frac{3}{4(\nu+p-3)} \sum_{j=2}^{p} x_j \\
&\quad + \frac{p+1}{2(\nu+p-3)} \sum_{j<k} x_j x_k, \\
E\,[Var\,(X_1 \mid \mathbf{X}_2)] &= \frac{\nu}{\nu-2} \frac{p+1}{2p}, \\
Var\,(\mathbf{X}_2) &= \frac{\nu}{\nu-2} \mathbf{R}_{22}, \\
Cov\,(X_1, X_i) &= -\frac{\nu}{p(\nu-2)}, \qquad i = 2, \ldots, p.
\end{aligned}
$$

Furthermore, the residual variance of X_1 with respect to $\mathbf{X}_2$ is

$$
\frac{\nu}{\nu-2} \frac{p+1}{2p},
$$

and the partial correlation coefficient of X_1 with respect to $\mathbf{X}_2$ is $-1/2$.

Patil and Kovner (1968) provided a detailed study of the trivariate t density

$$
\begin{aligned}
f\,(x_1, x_2, x_3) &= \frac{\Gamma\,((n+3)/2)}{(n\pi)^{3/2} \sqrt{1-\rho^2}\,\Gamma\,(n/2)} \\
&\quad \times \left(1 + \frac{1}{n} \frac{x_1^2 - 2\rho x_1 + x_2^2}{1-\rho^2} + x_3^2 \right)^{-(n+3)/2}.
\end{aligned}
$$

Among other results, Taylor series expansions – in powers of $1/n$ – of the density and associated probabilities in rectangles were given.