MULTIVARIATE T-NORMAL DISTRIBUTIONS IN APPLIED STATISTICS*

Yu. N. Blagoveshchenskii (Moscow, Russia) UDC 519.2

We study the family of so-called T-normal distributions, which includes the multivariate Gaussian distributions and those derived from them by arbitrary strictly monotone transformations of individual components. An attempt is made to give a quite thorough presentation of the theory of T-normal distributions and provide us with a series of appropriate statistical analysis procedures, together with some examples. In conclusion, various schemes of genesis of experimental data where the hypothesis about T-normality can be accepted are considered.

1. Introduction

This paper is devoted to the family of T-normal distributions, which contains all Gaussian distributions and all those distributions derived from them by arbitrary strictly monotone transformations of the components.

In many studies, fair attempts were made to transform the scales in order to normalize the distribution of individual characteristics. It was implicitly assumed that the joint distribution of the transformed characteristic became the multivariate normal distribution and the great body of available statistical procedures could be used.

Unfortunately, in the vast majority of cases the following is not taking into account:

(i) for any $p$-variate distribution with continuous and, say, bounded density, there always exist transformations $Y_i = H_i(X_i)$ of individual characteristics $X_1, \ldots, X_p$ such that the resulting variables $Y_i, i = 1, \ldots, p$, become Gaussian;

(ii) the fact that each of the individual variables $Y_1, \ldots, Y_p$ is distributed by the Gauss law does not imply that their joint distribution is the multivariate normal one.

There are many well-known examples for (ii). In particular, if the distribution of “new” characteristics $Y_1$ and $Y_2$ has the density

$$p(y_1, y_2) \equiv \frac{1}{2\pi} \exp \left\{ -\frac{y_1^2 + y_2^2}{2} \left( 1 + \frac{2y_1y_2}{y_1^2 + y_2^2} \right) \right\},$$

then each of the $Y_1$ and $Y_2$ is distributed by the standard Gauss law, although their joint distribution is not normal.

Assertion (i) obviously follows from the fact that for any random variable $\xi$ with continuous and strictly monotone distribution function $F(t)$ the variable $\eta \equiv F(\xi)$ is uniformly distributed on $[0; 1]$ (because the inequalities $\xi < t$ and $F(\xi) < F(t)$ are equivalent). Therefore, $\zeta \equiv \Phi^{-1}(F(\xi))$ has the standard Gauss distribution $\Phi(t)$ with density

$$\varphi(t) = (2\pi)^{-1/2} \exp \left\{ -\frac{t^2}{2} \right\}.$$

Thus, to validate (i) it suffices to set

$$H_i(t) \equiv \Phi^{-1}(F_i(t)), \quad F_i(t) \equiv P\{X_i < t\}, \quad i = 1, \ldots, p.$$  

Thus, transformations (1) of $X_1, \ldots, X_p$, provided that the density $f(x)$ of their joint distribution $F(x)$ in the $p$-dimensional Euclidean space $\mathbb{R}^p$ is continuous and bounded, allows us to pass to variables $Y_1, \ldots, Y_p$ whose univariate distributions are $P\{Y_i < t\} \equiv \Phi(t), i = 1, \ldots, p$. But this passage is of sense only in the case where the distribution $F(x)$ of $X_1, \ldots, X_p$ is $T$-normal (transformable to normal), i.e., by (1) the joint distribution of $Y_1, \ldots, Y_p$ is Gaussian.

Otherwise, we have no grounds for using the procedures for statistical analysis of Gaussian populations.

Some recommendations about scale normalizations have emerged in statistical studies many times, but they mainly concerned the search for convenient parametric families of such transformations (see [2] and [11], where curve linearization was considered).

It seems likely that the interest in the family of $T$-normal distributions was first pronounced in [10], but there this family did not emerge as a family in its own right. Only in [13] was this family presented together with formulas for the density and its estimation from a sample. Later (see [3]), algorithms to visualize multivariate data of the $T$-normal distribution were considered. Some elaborations of theoretical and applied character were contained in [4, 14].

*Supported by INTAS (grant No. 93-725-exp.).

Translated from Statisticheskie Metody Otsenivaniya i Proverki Gipotez, pp. 43–64, Perm, Russia, 1998.

556 1072-3374/01/1035-0556$25.00 © 2001 Plenum Publishing Corporation
2. The Family of T-Normal Distributions

2.1. Notation. Let \( a, b, x, y, \) and \( z \) denote \( p \)-dimensional vectors (columns), that is, points of the Euclidean space \( \mathbb{R}^p \). If \( x \in \mathbb{R}^p \), then it can be represented as a row-vector \( x^\top = (x_1, \ldots, x_p) \), where \( \top \) stands for transposition. The distance between points \( x \) and \( y \) is \( |x - y| \), where \( |x| = \sqrt{x^\top x} \). We write that a \( p \times p \) matrix \( B > 0 \) if it is positive definite; let \( y = a + Lx \), where \( L \) is an arbitrary \( p \times p \) matrix, denote a common linear transformation with linear operator \( L \) and shift \( a \in \mathbb{R}^p \).

We pass over the formal design of the statistical space and related problems. Generally speaking, the elementary event space \( \Omega \) is the set of \( \omega \in \mathbb{R}^p \), the \( \sigma \)-algebra of events \( \mathcal{A} \) is identified with the totality of the sets \( A \subseteq \mathbb{R}^p \) that are measured in the Lebesgue sense, and then the design presented in [9, Chap. 1] is used. If \( \xi = \xi(\omega) \) is a random vector with distribution \( F(x) \), then the probability of the event \( \{ \omega: \xi(\omega) \in A \} \) for any \( A \in \mathcal{A} \) is \( F(A) = \mathbb{P}\{\xi \in A\} \), where \( \mathbb{P}\{\cdot\} \) is the probability symbol.

Let the probability measure \( F(A) \) have density \( f(x) \). Then the marginal distributions \( F_j(t) = \mathbb{P}\{\xi_j < t\} \) of the random vector \( \xi \) have densities \( f_j(t) \), hence the quantiles \( F_j^{-1}(\lambda) \) are well defined with any \( 0 < \lambda < 1 \) as the root of the equation \( F_j(t) = \lambda \) in \( t, j = 1, \ldots, p \). For such a distribution \( F \), following (1), we introduce the transformation \( y = H(x) \) from \( \mathbb{R}^p \) into \( \mathbb{R}^p \), setting \( y_i = \Phi^{-1}(F_i(x_i)), \quad i = 1, 2, \ldots, p \).

As we have demonstrated, the random vector \( \eta = H(\xi) \) is the normalization of the vector \( \xi \); namely, all its components have the standard Gaussian distribution \( \Phi(t) \) with density

\[
\varphi(t) = (2\pi)^{-1/2} \exp\left\{ -\frac{t^2}{2} \right\}.
\]

We stress the fact that the joint distribution of the components of \( \eta \) is not necessarily Gaussian and for \( \eta \) the correlation matrix coincides with the covariance matrix \( C = E\eta\eta^\top \). Of course, if \( C = I \), where \( I \) is the \( p \times p \) identity matrix, then \( \eta \) has the standard \( p \)-variate Gaussian distribution \( \Phi_p(x) \) with density

\[
\varphi_p(x) = (2\pi)^{-p/2} \exp\left\{ -\frac{|x|^2}{2} \right\}.
\]

For a distribution \( F(x) \) with absolutely continuous marginal distributions \( F_i(t) \) and their densities

\[
f_i(t) = \frac{d}{dt} F_i(t),
\]
we introduce the function

\[
f^{(p)}(x) = \prod_{i=1}^p f_i(x_i).
\]

The distribution of the random vector \( \zeta = a + L\xi \), where \( \xi \) has the distribution \( \Phi_p(x), a \in \mathbb{R}^p \), and \( L \) is an arbitrary nonsingular \( p \times p \) matrix, is the \( p \)-variate normal distribution

\[
\Phi_p(x \mid a; C) = \frac{1}{(2\pi)^{p/2} \sqrt{|\det C|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left\{ -\frac{(y-a)^\top C^{-1}(y-a)}{2} \right\} dy,
\]

where \( a = E\xi \) is the mathematical expectation, \( C = E(\zeta - a)(\zeta - a)^\top = LL^\top \) is the covariance matrix, and \( \det C \) is the determinant of \( C \).

2.2. Distribution. A random vector \( \xi \) is said to be \( T \)-normal if the transformation \( \eta = H(\xi) \) is defined (its distribution \( F(x) \) in \( \mathbb{R}^p \) is absolutely continuous) and has the \( p \)-variate normal distribution. Moreover, for \( \eta = H(\xi) \) we have to set in (2) \( a = 0 \) and \( C = E\eta\eta^\top \), and the diagonal of the matrix \( C \) consists of ones, \( H_i(t) \equiv \Phi^{-1}(F_i(t)), \quad i \in \{1, \ldots, p\} \).