Aspects of Multivariate Statistical Theory

ROBB J. MUIRHEAD
Senior Statistical Scientist
Pfizer Global Research and Development
New London, Connecticut

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Aspects of Multivariate Statistical Theory

Aspects of Multivariate Statistical Theory

ROBB J. MUIRHEAD
Senior Statistical Scientist
Pfizer Global Research and Development
New London, Connecticut

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication is available.

ISBN-13 978-0-471-76985-9
ISBN-10 0-471-76985-1

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To
Nan and Bob
and
Maria and Mick

Preface

This book has grown out of lectures given in first- and second-year graduate courses at Yale University and the University of Michigan. It is designed as a text for graduate level courses in multivariate statistical analysis, and I hope that it may also prove to be useful as a reference book for research workers interested in this area.

Any person writing a book in multivariate analysis owes a great debt to T. W. Anderson for his 1958 text, *An Introduction to Multivariate Statistical Analysis*, which has become a classic in the field. This book synthesized various subareas for the first time in a broad overview of the subject and has influenced the direction of recent and current research in theoretical multivariate analysis. It is also largely responsible for the popularity of many of the multivariate techniques and procedures in common use today.

The current work builds on the foundation laid by Anderson in 1958 and in large part is intended to describe some of the developments that have taken place since then. One of the major developments has been the introduction of zonal polynomials and hypergeometric functions of matrix argument by A. T. James and A. G. Constantine. To a very large extent these have made possible a unified study of the noncentral distributions that arise in multivariate analysis under the standard assumptions of normal sampling. This work is intended to provide an introduction to some of this theory.

Most books of this nature reflect the author's tastes and interests, and this is no exception. The main focus of this work is on distribution theory, both exact and asymptotic. Multivariate techniques depend heavily on latent roots of random matrices; all of the important latent root distributions are introduced and approximations to them are discussed. In testing problems the primary emphasis here is on likelihood ratio tests and the distributions of likelihood ratio test statistics. The noncentral distributions are needed to evaluate power functions. Of course, in the absence of "best" tests simply computing power functions is of little interest; what is needed is a comparison of powers of competing tests over a wide range of alternatives. Wherever possible the results of such power studies in the literature are discussed. It should be mentioned, however, that although the emphasis is on likelihood ratio statistics, many of the techniques introduced here for studying and approximating their distributions can be applied to other test statistics as well.

A few words should be said about the material covered in the text. Matrix theory is used extensively, and matrix factorizations are extremely important. Most of the relevant material is reviewed in the Appendix, but some results also appear in the text and as exercises. Chapter 1 introduces the multivariate normal distribution and studies its properties, and also provides an introduction to spherical and elliptical distributions. These form an important class of non-normal distributions which have found increasing use in robustness studies where the aim is to determine how sensitive existing multivariate techniques are to multivariate normality assumptions. In Chapter 2 many of the Jacobians of transformations used in the text are derived, and a brief introduction to invariant measures via exterior differential forms is given. A review of matrix Kronecker or direct products is also included here. The reason this is given at this point rather than in the Appendix is that very few of the students that I have had in multivariate analysis courses have been familiar with this product, which is widely used in later work. Chapter 3 deals with the Wishart and multivariate beta distributions and their properties. Chapter 4, on decision-theoretic estimation of the parameters of a multivariate normal distribution, is rather an anomaly. I would have preferred to incorporate this topic in one of the other chapters, but there seemed to be no natural place for it. The material here is intended only as an introduction and certainly not as a review of the current state of the art. Indeed, only admissibility (or rather, inadmissibility) results are presented, and no mention is even made of Bayes procedures. Chapter 5 deals with ordinary, multiple, and partial correlation coefficients. An introduction to invariance theory and invariant tests is given in Chapter 6. It may be wondered why this topic is included here in view of the coverage of the relevant basic material in the books by E. L. Lehmann, *Testing Statistical Hypotheses*, and T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. The answer is that most of the students that have taken my multivariate analysis courses have been unfamiliar with invariance arguments, although they usually meet them in subsequent courses. For this reason I have long felt that an introduction to invariant tests in a multivariate text would certainly not be out of place.

Chapter 7 is where this book departs most significantly from others on multivariate statistical theory. Here the groundwork is laid for studying the noncentral distribution theory needed in subsequent chapters, where the emphasis is on testing problems in standard multivariate procedures. Zonal polynomials and hypergeometric functions of matrix argument are introduced, and many of their properties needed in later work are derived. Chapter 8 examines properties, and central and noncentral distributions, of likelihood ratio statistics used for testing standard hypotheses about covariance matrices and mean vectors. An attempt is also made here to explain what happens if these tests are used and the underlying distribution is non-normal. Chapter 9 deals with the procedure known as principal components, where much attention is focused on the latent roots of the sample covariance matrix. Asymptotic distributions of these roots are obtained and are used in various inference problems. Chapter 10 studies the multivariate general linear model and the distribution of latent roots and functions of them used for testing the general linear hypothesis. An introduction to discriminant analysis is also included here, although the coverage is rather brief. Finally, Chapter 11 deals with the problem of testing independence between a number of sets of variables and also with canonical correlation analysis.

The choice of the material covered is, of course, extremely subjective and limited by space requirements. There are areas that have not been mentioned and not everyone will agree with my choices; I do believe, however, that the topics included form the core of a reasonable course in classical multivariate analysis. Areas which are not covered in the text include factor analysis, multiple time series, multidimensional scaling, clustering, and discrete multivariate analysis. These topics have grown so large that there are now separate books devoted to each. The coverage of classification and discriminant analysis also is not very extensive, and no mention is made of Bayesian approaches; these topics have been treated in depth by Anderson and by Kshirsagar, *Multivariate Analysis*, and Srivastava and Khatri, *An Introduction to Multivariate Statistics*, and a person using the current work as a text may wish to supplement it with material from these references.

This book has been planned as a text for a two-semester course in multivariate statistical analysis. By an appropriate choice of topics it can also be used in a one-semester course. One possibility is to cover Chapters 1, 2, 3, 5, and possibly 6, and those sections of Chapters 8, 9, 10 and 11 which do not involve noncentral distributions and consequently do not utilize the theory developed in Chapter 7. The book is designed so that for the most part these sections can be easily identified and omitted if desired. Exercises are provided at the end of each chapter. Many of these deal with points which are alluded to in the text but left unproved. A few words are also in order concerning the Bibliography. I have not felt it necessary to cite the source of every result included here. Many of the original results due to such people as Wilks, Hotelling, Fisher, Bartlett, Wishart, and Roy have become so well known that they are now regarded as part of the folklore of multivariate analysis. T. W. Anderson's book provides an extensive bibliography of work prior to 1958, and my references to early work are indiscriminate at best. I have tried to be much more careful concerning references to the more recent work presented in this book, particularly in the area of distribution theory. No doubt some references have been missed, but I hope that the number of these is small. Problems which have been taken from the literature are for the most part not referenced unless the problem is especially complex or the reference itself develops interesting extensions and applications that the problem does not cover.

This book owes much to many people. My teachers, A. T. James and A. G. Constantine, have had a distinctive influence on me and their ideas are in evidence throughout, and especially in Chapters 2, 3, 7, 8, 9, 10, and 11. I am indebted to them both. Many colleagues and students have read, criticized, and corrected various versions of the manuscript. J. A. Hartigan read the first four chapters, and Paul Sampson used parts of the first nine chapters for a course at the University of Chicago; I am grateful to both for their extensive comments, corrections, and suggestions. Numerous others have also helped to weed out errors and have influenced the final version; especially deserving of thanks are D. Bancroft, W. J. Glynn, J. Kim, M. Kramer, R. Kuick, D. Marker, and J. Wagner. It goes without saying that the responsibility for all remaining errors is mine alone. I would greatly appreciate being informed about any that are found, large and small.

A number of people tackled the unenviable task of typing various parts and revisions of the manuscript. For their excellent work and their patience with my handwriting I would like to thank Carol Hotton, Terri Lomax Hunter, Kelly Kane, and Deborah Swartz.

ROBB J. MUIRHEAD

Ann Arbor, Michigan
February 1982

Contents

Tables

Commonly Used Notation

R^m	Euclidean space of dimension m consisting of $m \times 1$ real vectors
det	determinant
S_m	unit sphere in R^m centered at the origin
$\wedge$	exterior or wedge product
Re	real part
tr	trace
etr	exp tr
$A > 0$	A is positive definite
$V_{m,n}$	Stiefel manifold of $n \times m$ matrices with orthonormal columns
$O(m)$	Group of orthogonal $m \times m$ matrices
$\otimes$	direct or Kronecker product
$\lVert \mathbf{X} \rVert$	norm of $\mathbf{X}$ i.e. $\left(\sum_{i=1}^{m} X_i^2 \right)^{\frac{1}{2}}$
$\mathcal{S}_m$	set of all $m \times m$ positive definite matrices
$\mathcal{Gl}(m, R)$	general linear group of $m \times m$ nonsingular real matrices
$\mathcal{Al}(m, R)$	affine group $\{(B, \mathbf{c}); B \in \mathcal{Gl}(m, R); \mathbf{c} \in R^m\}$
$C_\kappa(\cdot)$	zonal polynomial
$M_\kappa(\cdot)$	monomial symmetric function
$(\alpha)_\kappa$	generalized hypergeometric coefficient
$\lVert X \rVert$	max of absolute values of latent roots of the matrix X
$\binom{\kappa}{\sigma}$	generalized binomial coefficient
$L_\kappa^\gamma(\cdot)$	generalized Laguerre polynomial
$N_m(\boldsymbol{\mu}, \Sigma)$	m-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ
$W_m(n, \Sigma)$	$m \times m$ matrix variate Wishart distribution with n degrees of freedom and covariance matrix Σ
$\mathrm{Beta}_m(\alpha, \beta)$	$m \times m$ matrix variate beta distribution with parameters α, β

Aspects of Multivariate Statistical Theory

CHAPTER 1

The Multivariate Normal and Related Distributions

1.1. INTRODUCTION

The basic, central distribution and building block in classical multivariate analysis is the multivariate normal distribution. There are two main reasons why this is so. First, it is often the case that multivariate observations are, at least approximately, normally distributed. This is especially true of sample means and covariance matrices used in formal inferential procedures, due to a *central limit theorem* effect. This effect is also felt, of course, when the observations themselves can be regarded as sums of independent random vectors or effects, a realistic model in many situations. Secondly, the multivariate normal distribution and the sampling distributions it gives rise to are, in the main, tractable. This is not generally the case with other multivariate distributions, even for ones which appear to be close to the normal.

We will be concerned primarily with classical multivariate analysis, that is, techniques, distributions, and inferences based on the multivariate normal distribution. This distribution is defined in Section 1.2 and various properties are also derived there. This is followed by a review of the noncentral χ^2 and F distributions in Section 1.3 and some results about quadratic forms in normal variables in Section 1.4.

A natural question is to ask what happens to the inferences we make under the assumption of normality if the observations are not normal. This is an important question, leading into the area that has come to be known generally as robustness. In Section 1.5 we introduce the class of elliptical distributions; these distributions have been commonly used as alternative models in robustness studies. Section 1.6 reviews some results about multivariate cumulants. For our purposes, these are important in asymptotic distributions of test statistics which are functions of a sample covariance matrix.

It is expected that the reader is familiar with basic distributions such as the normal, gamma, beta, t, and F and with the concepts of jointly distributed random variables, marginal distributions, moments, conditional distributions, independence, and related topics covered in such standard probability and statistics texts as Bickel and Doksum (1977) and Roussas (1973).

Characteristic functions and basic limit theorems are also important and useful references are Cramér (1946), Feller (1971), and Rao (1973). Matrix notation and theory is used extensively; some of this theory appears in the text and some is reviewed in the Appendix.

1.2. THE MULTIVARIATE NORMAL DISTRIBUTION

1.2.1. Definition and Properties

Before proceeding to the multivariate normal distribution we need to define some moments of a *random* vector, i.e., a vector whose components are jointly distributed. The *mean* or *expectation* of a random $m\times 1$ vector $\mathbf{X}=(X_1,\ldots,X_m)'$ is defined to be the vector of expectations:

$$E(\mathbf{X})=\begin{pmatrix} E(X_1)\\ \vdots \\ E(X_m)\end{pmatrix}.$$

More generally, if $Z=(z_{ij})$ is a $p\times q$ random matrix then $E(Z)$, the expectation of Z, is the matrix whose $i-j$th element is $E(z_{ij})$. It is a simple matter to check that if B, C and D are $m\times p$, $q\times n$ and $m\times n$ matrices of constants, then

$$E(BZC+D)=BE(Z)C+D. \tag{1}$$

If $\mathbf{X}$ has mean $\boldsymbol{\mu}$ the covariance matrix of $\mathbf{X}$ is defined to be the $m\times m$ matrix

$$\Sigma\equiv\operatorname{Cov}(\mathbf{X})=E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right].$$

The $i-j$th element of Σ is

$$\sigma_{ij}=E\left[(X_i-\mu_i)(X_j-\mu_j)\right],$$

the covariance between X_i and X_j, and the $i-i$th element is

$$\sigma_{ii} = E\left[(X_i - \mu_i)^2\right],$$

the variance of X_i, so that the diagonal elements of Σ must be nonnegative. Obviously Σ is symmetric, i.e., $\Sigma = \Sigma'$. Indeed, the class of covariance matrices coincides with the class of *non-negative definite* matrices. Recall that an $m \times m$ symmetric matrix A is called non-negative definite if

$$\boldsymbol{\alpha}' A \boldsymbol{\alpha} \geq 0 \qquad \text{for all} \quad \boldsymbol{\alpha} \in R^m$$

and positive definite if

$$\boldsymbol{\alpha}' A \boldsymbol{\alpha} > 0 \qquad \text{for all} \quad \boldsymbol{\alpha} \in R^m, \quad \boldsymbol{\alpha} \neq \mathbf{0}$$

(Here, and throughout the book, R^m denotes Euclidean space of m dimensions consisting of $m \times 1$ vectors with real components.)

LEMMA 1.2.1. The $m \times m$ matrix Σ is a covariance matrix if and only if it is non-negative definite.

Proof. Suppose Σ is the covariance matrix of a random vector $\mathbf{X}$, where $\mathbf{X}$ has mean $\boldsymbol{\mu}$, Then for all $\boldsymbol{\alpha} \in R^m$,

$$\begin{aligned} \text{(2)} \qquad \operatorname{Var}(\boldsymbol{\alpha}'\mathbf{X}) &= E\left[(\boldsymbol{\alpha}'\mathbf{X} - \boldsymbol{\alpha}'\boldsymbol{\mu})^2\right] \\ &= E\left[(\boldsymbol{\alpha}'(\mathbf{X} - \boldsymbol{\mu}))^2\right] \\ &= E\left[\boldsymbol{\alpha}'(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\alpha}\right] \\ &= \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha} \geq 0 \end{aligned}$$

so that Σ is non-negative definite. Now suppose Σ is a non-negative definite matrix of rank r, say $(r \leq m)$. Write $\Sigma = CC'$, where C is an $m \times r$ matrix of rank r (see Theorem A9.4). Let $\mathbf{Y}$ be an $r \times 1$ vector of independent random variables with mean $\mathbf{0}$ and $\operatorname{Cov}(\mathbf{Y}) = I$ and put $\mathbf{X} = C\mathbf{Y}$. Then $E(\mathbf{X}) = \mathbf{0}$ and

$$\begin{aligned} \operatorname{Cov}(\mathbf{X}) = E[\mathbf{X}\mathbf{X}'] &= E[C\mathbf{Y}\mathbf{Y}'C'] \\ &= CE(\mathbf{Y}\mathbf{Y}')C' \\ &= CC' = \Sigma, \end{aligned}$$

so that Σ is a covariance matrix.

As a direct consequence of the inequality (2) we see that if the covariance matrix Σ of a random vector $\mathbf{X}$ is not positive definite then, with probability 1, the components X_i of $\mathbf{X}$ are linearly related. For then there exists $\boldsymbol{\alpha} \in R^m$, $\boldsymbol{\alpha} \neq \mathbf{0}$, such that

$$\operatorname{Var}(\boldsymbol{\alpha}'\mathbf{X}) = \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha} = 0$$

so that, with probability 1, $\boldsymbol{\alpha}'\mathbf{X} = k$, where $k = \boldsymbol{\alpha}'E(\mathbf{X})$—which means that $\mathbf{X}$ lies in a hyperplane.

We will commonly make linear transformations of random vectors and will need to know how covariance matrices are transformed. Suppose $\mathbf{X}$ is an $m \times 1$ random vector with mean $\boldsymbol{\mu}_x$ and covariance matrix Σ_x and let $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, where B is $k \times m$ and $\mathbf{b}$ is $k \times 1$. The mean of $\mathbf{Y}$ is, by (1), $\boldsymbol{\mu}_y = B\boldsymbol{\mu}_x + \mathbf{b}$, and the covariance matrix of $\mathbf{Y}$ is

$$\begin{aligned}
(3) \qquad \Sigma_y &= E\left[(Y-\boldsymbol{\mu}_y)(Y-\boldsymbol{\mu}_y)'\right] \\
&= E\left[(B\mathbf{X}+\mathbf{b}-(B\boldsymbol{\mu}_x+\mathbf{b}))(B\mathbf{X}+\mathbf{b}-(B\boldsymbol{\mu}_x+\mathbf{b}))'\right] \\
&= BE\left[(\mathbf{X}-\boldsymbol{\mu}_x)(\mathbf{X}-\boldsymbol{\mu}_x)'\right]B' \\
&= B\Sigma_x B'.
\end{aligned}$$

In order to define the multivariate normal distribution we will use the following result.

THEOREM 1.2.2. If $\mathbf{X}$ is an $m \times 1$ random vector then its distribution is uniquely determined by the distributions of linear functions $\boldsymbol{\alpha}'\mathbf{X}$, for every $\boldsymbol{\alpha} \in R^m$.

Proof. The characteristic function of $\boldsymbol{\alpha}'\mathbf{X}$ is

$$\phi(t, \boldsymbol{\alpha}) = E\left[e^{it\boldsymbol{\alpha}'\mathbf{X}}\right]$$

so that

$$\phi(1, \boldsymbol{\alpha}) = E\left[e^{i\boldsymbol{\alpha}'\mathbf{X}}\right],$$

which, considered as a function of $\boldsymbol{\alpha}$, is the characteristic function of $\mathbf{X}$ (i.e., the joint characteristic function of the components of $\mathbf{X}$). The required result then follows by invoking the fact that a distribution in R^m is uniquely determined by its characteristic function [see, e.g., Cramér (1946), Section 10.6, or Feller (1971), Section XV.7].

DEFINITION 1.2.3. The $m\times 1$ random vector $\mathbf{X}$ is said to have an m-variate normal distribution if, for every $\boldsymbol{\alpha}\in R^m$, the distribution of $\boldsymbol{\alpha}'\mathbf{X}$ is univariate normal.

Proceeding from this definition we will now establish some properties of the multivariate normal distribution.

THEOREM 1.2.4. If $\mathbf{X}$ has an m-variate normal distribution then both $\boldsymbol{\mu}\equiv E(\mathbf{X})$ and $\Sigma\equiv\mathrm{Cov}(\mathbf{X})$ exist and the distribution of $\mathbf{X}$ is determined by $\boldsymbol{\mu}$ and Σ.

Proof. If $\mathbf{X}=(X_1,\dots,X_m)'$ then, for each $i=1,\dots,m$, X_i is univariate normal (using Definition 1.2.3) so that $E(X_i)$ and $\mathrm{Var}(X_i)$ exist and are finite. Thus $\mathrm{Cov}(X_i, X_j)$ exists. (Why?) Putting $\boldsymbol{\mu}=E(\mathbf{X})$ and $\Sigma=\mathrm{Cov}(\mathbf{X})$, we have, from (1) and (3),

$$E(\boldsymbol{\alpha}'\mathbf{X})=\boldsymbol{\alpha}'\boldsymbol{\mu}$$

and

$$\mathrm{Var}(\boldsymbol{\alpha}'\mathbf{X})=\boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}$$

so that the distribution of $\boldsymbol{\alpha}'\mathbf{X}$ is $N(\boldsymbol{\alpha}'\boldsymbol{\mu}, \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha})$ for each $\boldsymbol{\alpha}\in R^m$. Since these univariate distributions are determined by $\boldsymbol{\mu}$ and Σ so is the distribution of $\mathbf{X}$ by Theorem 1.2.2.

The m-variate normal distribution of the random vector $\mathbf{X}$ of Theorem 1.2.4 will be denoted by $N_m(\boldsymbol{\mu},\Sigma)$ and we will write that $\mathbf{X}$ is $N_m(\boldsymbol{\mu},\Sigma)$.

THEOREM 1.2.5. If $\mathbf{X}$ is $N_m(\boldsymbol{\mu},\Sigma)$ then the characteristic function of $\mathbf{X}$ is

$$\phi_{\mathbf{x}}(\mathbf{t})=\exp\left(i\boldsymbol{\mu}'\mathbf{t}-\tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right). \tag{4}$$

Proof. Here

$$\phi_{\mathbf{x}}(\mathbf{t})=E\left[e^{i\mathbf{t}'\mathbf{X}}\right]=\phi_{\mathbf{t}'\mathbf{x}}(1),$$

where the right side denotes the characteristic function of the random variable $\mathbf{t}'\mathbf{X}$ evaluated at 1. Since $\mathbf{X}$ is $N_m(\boldsymbol{\mu},\Sigma)$ then $\mathbf{t}'\mathbf{X}$ is $N(\mathbf{t}'\boldsymbol{\mu},\mathbf{t}'\Sigma\mathbf{t})$ so that

$$\phi_{\mathbf{t}'\mathbf{x}}(1)=\exp\left(i\mathbf{t}'\boldsymbol{\mu}-\tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right),$$

completing the proof.

The alert reader may have noticed that we have not yet established the existence of the multivariate normal distribution. It could be that Definition 1.2.3 is vacuous! To sew things up we will show that the function given by (4) is indeed the characteristic function of a random vector. Let Σ be an $m\times m$ covariance matrix (i.e., a non-negative definite matrix) of rank r and let $U_1,\dots,U_r$ be independent standard normal random variables. The vector $\mathbf{U}=(U_1,\dots,U_r)'$ has characteristic function

$$\begin{aligned}\phi_{\mathbf{u}}(\mathbf{t}) &= E[\exp(i\mathbf{t}'\mathbf{U})]\\ &= \prod_{j=1}^{r} E[\exp(it_jU_j)] \qquad \text{(by independence)}\\ &= \prod_{j=1}^{r} \exp(-\tfrac{1}{2}t_j^2) \qquad \text{(by normality)}\\ &= \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{t}).\end{aligned}$$

Now put

$$\mathbf{X}=C\mathbf{U}+\boldsymbol{\mu} \tag{5}$$

where C is an $m\times r$ matrix of rank r such that $\Sigma=CC'$, and $\boldsymbol{\mu}\in R^m$. Then $\mathbf{X}$ has characteristic function (4), for

$$\begin{aligned}E[\exp(i\mathbf{t}'\mathbf{X})] &= E[\exp(i\mathbf{t}'C\mathbf{U})]\exp(i\mathbf{t}'\boldsymbol{\mu})\\ &= \phi_{\mathbf{u}}(C'\mathbf{t})\exp(i\mathbf{t}'\boldsymbol{\mu})\\ &= \exp(-\tfrac{1}{2}\mathbf{t}'CC'\mathbf{t})\exp(i\boldsymbol{\mu}'\mathbf{t})\\ &= \exp(i\boldsymbol{\mu}'\mathbf{t}-\tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}).\end{aligned}$$

It is worth remarking that we could have *defined* the multivariate normal distribution $N_m(\boldsymbol{\mu},\Sigma)$ by means of the linear transformation (5) on independent standard normal variables. Such a representation is often useful; see, for example, the proof of Theorem 1.2.9.

Getting back to the properties of the multivariate normal distribution our next result shows that any linear transformation of a normal vector has a normal distribution.

THEOREM 1.2.6. If $\mathbf{X}$ is $N_m(\boldsymbol{\mu},\Sigma)$ and B is $k\times m$, $\mathbf{b}$ is $k\times 1$ then

$$\mathbf{Y}=B\mathbf{X}+\mathbf{b} \quad \text{is} \quad N_k(B\boldsymbol{\mu}+\mathbf{b},\, B\Sigma B').$$