Heliyon

ELSEVIER

Received:
8 April 2018
Revised:
4 August 2018
Accepted:
10 January 2019

Cite as:
Rajeshwari Majumdar, Suman Majumdar. On the conditional distribution of a multivariate Normal given a transformation – the linear case.
Heliyon 5 (2019) e01136.
doi: 10.1016/j.heliyon.2019.e01136

On the conditional distribution of a multivariate Normal given a transformation – the linear case

Rajeshwari Majumdar [a,*], **Suman Majumdar** [b]

[a] *Department of Politics, New York University, 19 West 4th Street, New York, NY 10012, United States of America*

[b] *Department of Statistics, University of Connecticut, 1 University Place, Stamford, CT 06901, United States of America*

* Corresponding author.
E-mail address: majumdar@nyu.edu (R. Majumdar).

Abstract

We show that the orthogonal projection operator onto the range of the adjoint T^* of a linear operator T can be represented as UT, where U is an invertible linear operator. Given a Normal random vector Y and a linear operator T, we use this representation to obtain a linear operator $\widehat{T}$ such that $\widehat{T}Y$ is independent of TY and $Y - \widehat{T}Y$ is an affine function of TY. We then use this decomposition to prove that the conditional distribution of a Normal random vector Y given $\mathcal{T}Y$, where $\mathcal{T}$ is a linear transformation, is again a multivariate Normal distribution. This result is equivalent to the well-known result that given a k-dimensional component of a n-dimensional Normal random vector, where $k < n$, the conditional distribution of the remaining $(n-k)$-dimensional component is a $(n-k)$-dimensional multivariate Normal distribution, and sets the stage for approximating the conditional distribution of Y given $g(Y)$, where g is a continuously differentiable vector field.

Keywords: Mathematics

1. Introduction

What can we ascertain about the conditional distribution of a multivariate Normal random vector $Y \in \mathbb{R}^n$ given $g(Y)$, where $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a measurable function?

https://doi.org/10.1016/j.heliyon.2019.e01136

Clearly, given a particular functional form of g, the problem is a very specific one, and depending on the functional form, may or may not have a closed form solution. Our objective is to derive an approximation to the conditional distribution in question based on some regularity properties of g. Specifically, in this paper we find the conditional distribution when g is a linear transformation; in a companion paper, we expect to derive the desired approximation when g is a continuously differentiable vector field, that is, an element of C^1, by exploiting the local linearity of g and the results of this paper.

Before proceeding further, we present a brief review of what is known about the conditional distribution when g is a linear transformation, to be denoted by $\mathcal{T}$ in what follows. Casella and Berger (2002, Definition 4.5.10) define the bivariate Normal distribution by specifying the joint density in terms of the five parameters of the distribution – the means μ_1 and μ_2, the variances σ_1^2 and σ_2^2, and the correlation ρ. They calculate the marginal density of Y_1 and note that it is easy to verify, using the joint and marginal densities, that the conditional distribution of Y_2 given $Y_1 = y_1$ is Normal with

$$\text{mean} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}\left(y_1 - \mu_1\right) \text{ and variance} = \sigma_2^2\left(1 - \rho^2\right). \tag{1}$$

Anderson (1984, Section 2.5.1) and Flury (1997, Theorem 3.3.1) generalize this result to the multivariate Normal distribution. While Anderson (1984) deals with the Lebesgue density of the multivariate Normal distribution (which exists only when the covariance matrix is of full rank), Flury (1997) avoids dealing with the density by defining the multivariate Normal distribution in terms of linear functionals, but requires the covariance matrix of the conditioning component to be of full rank.

Though their approaches to defining the multivariate Normal distribution differ, both Muirhead (1982, Theorem 1.2.11) and Eaton (1983, Proposition 3.13) obtain, as described below, the conditional distribution of one component of a multivariate Normal random vector given another component without any restriction on the rank of the covariance matrix of the conditioning component. Let $Y_{n\times 1}$ have the multivariate Normal distribution with mean vector $\mu_{n\times 1}$ and covariance matrix $\Sigma_{n\times n}$. Partition Y, μ, and Σ respectively as

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},\ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},\ \text{and } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{2}$$

where Y_1 and μ_1 are $k \times 1$ and Σ_{11} is $k \times k$. Then $Y_2 - \Sigma_{21}\Sigma_{11}^{-}Y_1$ and Y_1 are jointly Normal and uncorrelated, hence independent, where Σ_{11}^{-} is a generalized inverse of Σ_{11}. Consequently, the conditional distribution of $Y_2 - \Sigma_{21}\Sigma_{11}^{-}Y_1$ given Y_1 equals the unconditional distribution of $Y_2 - \Sigma_{21}\Sigma_{11}^{-}Y_1$, which is multivariate Normal in $(n-k)$ dimensions with mean $\mu_2 - \Sigma_{21}\Sigma_{11}^{-}\mu_1$ and covariance $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{12}$. Thus, the

 https://doi.org/10.1016/j.heliyon.2019.e01136

conditional distribution of Y_2 given Y_1 is multivariate Normal in $(n-k)$ dimensions with mean $\mu_2 + \Sigma_{21}\Sigma_{11}^{-}\left(Y_1 - \mu_1\right)$ and covariance $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{12}$.

Given two topological vector spaces V and W, let $\mathcal{L}(V, W)$ denote the linear space of continuous linear transformations from V to W, and let $\mathcal{L}(V, V) = \mathcal{L}(V)$. Recall from Muirhead (1982, Theorem 1.2.6) that if $\mathcal{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, the joint distribution of $(\mathcal{T}Y, Y)' \in \mathbb{R}^{m+n}$ is multivariate Normal with

$$\text{mean} = \begin{pmatrix} A\mu \\ \mu \end{pmatrix} \text{ and covariance} = \begin{bmatrix} A\Sigma A' & A\Sigma \\ \Sigma A' & \Sigma \end{bmatrix}, \tag{3}$$

where the $m \times n$ matrix A represents the transformation $\mathcal{T}$ with respect to the standard orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. By the results of Muirhead (1982) and Eaton (1983) cited in the previous paragraph, we obtain that the conditional distribution of Y given $\mathcal{T}Y$ is multivariate Normal in $\mathbb{R}^n$ with mean $\mu + \Sigma A'\left(A\Sigma A'\right)^{-} A(Y - \mu)$ and covariance $\Sigma - \Sigma A'\left(A\Sigma A'\right)^{-} A\Sigma$.

Unfortunately, this derivation of the conditional distribution of Y given $\mathcal{T}Y$, because of its dependence on manipulative matrix algebra, is not of much help when it comes to exploiting the local linearity of g for approximating the conditional distribution of Y given $g(Y)$ for a g in C^1. In what follows, we present an alternative derivation of the conditional distribution of Y given $\mathcal{T}Y$. Our derivation is expected to facilitate the approximation of the conditional distribution when the transformation g is nonlinear but continuously differentiable; see Remark 5.

We define a multivariate Normal distribution in terms of an affine transformation of the random vector with coordinates independent and identically distributed (iid, hereinafter) standard Normals, as in Muirhead (1982), but work with the covariance operator (instead of matrix) and the characteristic function. This coordinate-free approach allows us to seamlessly subsume the possibility that the multivariate Normal distribution is supported on a proper subspace of $\mathbb{R}^n$ which is not spanned by a subset of the standard orthonormal basis $\{e_1, \cdots, e_n\}$. In the spirit of Axler (2015), this relegates the manipulative matrix algebra that so dominates the multivariate Normal literature to the back burner.

In Theorem 3, given $T \in \mathcal{L}(\mathbb{R}^n)$ we find $\widehat{T} \in \mathcal{L}(\mathbb{R}^n)$, depending on T and the covariance of Y, such that $\widehat{T}Y$ is independent of TY and $Y - \widehat{T}Y$ is an affine function of TY. In Theorem 4, given $\mathcal{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ we find $T \in \mathcal{L}(\mathbb{R}^n)$ such that the conditional distribution of Y given TY equals that given $\mathcal{T}Y$, and use the decomposition obtained in Theorem 3 to obtain the conditional distribution of Y given TY, hence that given $\mathcal{T}Y$. Note that, since a component of a vector is a linear transformation of the vector and a linear transformation of a multivariate Normal random variable is another multivariate Normal random variable (Lemma 5),

Theorem 4 allows us to deduce Theorem 1.2.11(b) of Muirhead (1982) (or, for that matter, Proposition 3.13 of Eaton, 1983) as an immediate corollary.

The paper is organized as follows. In Section 2, we introduce all the background results from linear algebra used in Section 3. In Section 3, we present all the results on the multivariate Normal distribution, including Theorem 3 and Theorem 4.

2. Background

Let V, W be finite-dimensional real vector spaces. Since every finite-dimensional real vector space is isomorphic to an Euclidean space (Axler, 2015, Theorem 3.59), we assume, without loss of generality, that V and W are real inner product spaces. For any $S \in \mathcal{L}(V, W)$, let $\mathcal{R}(S) \subseteq W$ denote the range of S and $\mathcal{N}(S) \subseteq V$ denote the null space of S. For $T \in \mathcal{L}(V, W)$, let $T^* \in \mathcal{L}(W, V)$ denote the adjoint of T (Axler, 2015, Definition 7.2). For any subspace Q of V, let $Q^\perp$ denote the orthogonal complement of Q (Axler, 2015, Definition 6.45) and $\Pi_Q \in \mathcal{L}(V)$ denote the orthogonal projection operator onto Q (Axler, 2015, Definition 6.53).

The main result of this section is Theorem 1, where we show, using the Spectral Theorem, that the orthogonal projection operator onto the range of the adjoint T^* of a linear operator T is UT, where U is an invertible linear operator. The result of Lemma 1 is used in the proof of Theorem 1. In Lemma 2, using the Spectral Theorem again, we define the Moore–Penrose inverse and study some of its properties.

We first present Lemma 1, which is mentioned in Exercise 3.D.3 of Axler (2015); we sketch a proof here for the sake of completeness.

Lemma 1. *Let W be a subspace of V and $\mathcal{U} \in \mathcal{L}(W, V)$. There exists an invertible operator $U \in \mathcal{L}(V)$ such that $Uw = \mathcal{U}w$ for every $w \in W$ if and only if $\mathcal{U}$ is injective.*

Proof of Lemma 1. Let $U \in \mathcal{L}(V)$ be invertible such that $Uw = \mathcal{U}w$ for every $w \in W$. Let $w \in W$ be such that $\mathcal{U}w = 0$, implying $Uw = 0$. Since U is invertible, hence injective (Axler, 2015, Theorem 3.69), we obtain $w = 0$, showing that $\mathcal{U}$ is injective.

To show the converse, let Q be a direct sum complement of W and X a direct sum complement of $\mathcal{R}(\mathcal{U})$, where the existence of Q and X are guaranteed by Theorem 2.34 of Axler (2015). Let $\{q_1, \cdots, q_k\}$ be a basis for Q and $\{x_1, \cdots, x_m\}$ a basis for X. By the fundamental theorem of linear maps (Axler, 2015, Theorem 3.22), $\dim \mathcal{R}(\mathcal{U}) \leq \dim W$, implying, by Theorem 2.43 of Axler (2015), $k \leq m$. Recall that every $v \in V$ can be uniquely decomposed as $w + q$, where $w \in W$ and $q = \sum_{j=1}^{k} c_j q_j \in Q$. Define $U \in \mathcal{L}(V)$ as

$$Uv = \mathcal{U}w + x, \text{ where } x = \sum_{j=1}^{k} c_j x_j \in X. \tag{4}$$

Since X is a direct sum complement of $\mathcal{R}(\mathcal{U})$, $Uv = 0$ implies $\mathcal{U}w = x = 0$. Since $\mathcal{U}$ is injective and $\{x_1, \cdots, x_k\}$ is linearly independent, U is injective, hence invertible (again by Theorem 3.69 of Axler, 2015). □

Theorem 1. *Given $T \in \mathcal{L}(V)$, there exists $U \in \mathcal{L}(V)$, invertible and depending on T, such that $\Pi_{\mathcal{R}(T^*)} = UT$.*

Proof of Theorem 1. We first observe that T^*T is a positive operator (Axler, 2015, Definition 7.31). By Theorems 7.6(e) and 7.6(c) of Axler (2015), T^*T is self-adjoint; since, by the definition of the adjoint operator, for every $v \in V$,

$$\langle T^*Tv, v\rangle = \|Tv\|^2 \geq 0, \tag{5}$$

the positivity of T^*T follows.

By the Real Spectral Theorem (Axler, 2015, Theorem 7.29(b)), V has an orthonormal basis $\{f_1, \cdots, f_n\}$ consisting of eigenvectors of T^*T with corresponding eigenvalues $\{\lambda_1, \cdots, \lambda_n\}$. By Theorem 7.35(b) of Axler (2015), $\lambda_j \geq 0$ for all $1 \leq j \leq n$. Since, by (5), $\lambda_j = \left\|Tf_j\right\|^2$, we obtain

$$\begin{aligned} &\lambda_j = 0 \Leftrightarrow Tf_j = 0 \\ &\lambda_j > 0 \Leftrightarrow Tf_j \neq 0. \end{aligned} \tag{6}$$

If $\lambda_j = 0$ for every $j = 1, \cdots, n$, then T^*T is the zero operator, implying, by (5), that T is the zero operator. By Theorem 7.7 of Axler (2015), T^* is the zero operator as well, and the theorem trivially holds with $U = I$.

Thus, for the remainder of the proof, we assume that $\mathfrak{P} = \{j : 1 \leq j \leq n,\ \lambda_j > 0\}$ is non-empty. Let $\mathfrak{P}^c = \{j : 1 \leq j \leq n,\ \lambda_j = 0\}$.

For $j \in \mathfrak{P}$, since $T^*\left(\frac{Tf_j}{\lambda_j}\right) = f_j$, we have $f_j \in \mathcal{R}(T^*)$; that is,

$$\Pi_{\mathcal{R}(T^*)} f_j = f_j. \tag{7}$$

For $j \in \mathfrak{P}^c$, $f_j \in \mathcal{N}(T)$ by (6); since $\mathcal{N}(T) = (\mathcal{R}(T^*))^{\perp}$ (Axler, 2015, Theorem 7.7(c)), we have

$$\Pi_{\mathcal{R}(T^*)} f_j = 0. \tag{8}$$

By (7) and (8), for any $x \in V$,

$$\Pi_{\mathcal{R}(T^*)} x = \sum_{j \in \mathfrak{P}} \langle x, f_j\rangle f_j. \tag{9}$$

By (6) and the definition of $\mathfrak{P}^c$, for any $x \in V$,

$$Tx = \sum_{j \in \mathfrak{P}} \langle x, f_j \rangle T f_j. \tag{10}$$

By definition of f_j and λ_j, $1 \leq j \leq n$, the list of vectors $\left\{ Tf_j / \left\| Tf_j \right\| : j \in \mathfrak{P} \right\}$ in V is orthonormal, and consequently, by Theorem 6.26 of Axler (2015), linearly independent; the same conclusion holds for the list $\left\{ f_j : j \in \mathfrak{P} \right\}$. For $W = \operatorname{span} \left\{ Tf_j : j \in \mathfrak{P} \right\}$, $\mathcal{U} \in \mathcal{L}(W, V)$ defined by $\mathcal{U} T f_j = f_j$ is clearly injective. By Lemma 1 there exists an invertible operator $U \in \mathcal{L}(V)$ such that

$$Uw = \mathcal{U} w \text{ for every } w \in W. \tag{11}$$

Now note that, for any $x \in V$, by (10), (11), the definition of $\mathcal{U}$, and (9), in that order,

$$UTx = \sum_{j \in \mathfrak{P}} \langle x, f_j \rangle U T f_j = \sum_{j \in \mathfrak{P}} \langle x, f_j \rangle \mathcal{U} T f_j = \sum_{j \in \mathfrak{P}} \langle x, f_j \rangle f_j = \Pi_{\mathcal{R}(T^*)} x, \tag{12}$$

completing the proof. □

Lemma 2. *Given $D \in \mathcal{L}(V)$ positive, there exists $D^{-1/2} \in \mathcal{L}(V)$ positive such that*

$$D^{1/2} D^{-1/2} = D^{-1/2} D^{1/2} = \Pi_{\mathcal{R}(D)} \tag{13}$$

and

$$D D^{-1/2} = D^{1/2} = D^{-1/2} D, \tag{14}$$

where $D^{1/2}$ denotes the unique positive square root of D.

Remark 1. The operator $D^{-1/2}$ is the Moore–Penrose inverse (Penrose, 1955) of the operator $D^{1/2}$.

Proof of Lemma 2. By Theorem 7.36 of Axler (2015), $D^{1/2}$ exists and is defined by

$$D^{1/2} f_j = \lambda_j^{1/2} f_j, \tag{15}$$

where $\left\{ f_1, f_2, \cdots, f_n \right\}$ is an orthonormal basis of V consisting of eigenvectors of D with corresponding non-negative eigenvalues $\left\{ \lambda_1, \lambda_2, \cdots, \lambda_n \right\}$. Let

$$\mathfrak{P} = \left\{ j : 1 \leq j \leq n,\ \lambda_j > 0 \right\} \text{ and } \mathfrak{P}^c = \left\{ j : 1 \leq j \leq n,\ \lambda_j = 0 \right\}. \tag{16}$$

Let $D^{-1/2} \in \mathcal{L}(V)$ be defined by

$$D^{-1/2} f_j = \begin{cases} \lambda_j^{-1/2} f_j & \text{if } j \in \mathfrak{P} \\ 0 & \text{if } j \in \mathfrak{P}^c. \end{cases} \tag{17}$$

Clearly, $\left\langle D^{-1/2}x, y\right\rangle = \sum_{j\in\mathfrak{P}} \lambda_j^{-1/2}\langle x, f_j\rangle\langle y, f_j\rangle = \left\langle x, D^{-1/2}y\right\rangle$, showing that $D^{-1/2}$ is self-adjoint and implying $\left\langle D^{-1/2}x, x\right\rangle \geq 0$, that is, $D^{-1/2} \in \mathcal{L}(V)$ is positive.

For any $y \in V$, by (15) and (17),

$$D^{1/2}D^{-1/2}y = \sum_{j\in\mathfrak{P}} \langle y, f_j\rangle f_j = D^{-1/2}D^{1/2}y. \tag{18}$$

Since D is self-adjoint,

$$\mathcal{R}(D) = \left(\mathcal{N}(D)\right)^{\perp} \tag{19}$$

by Theorem 7.7(d) of Axler (2015); by Theorem 6.50 of Axler (2015),

$$\dim \mathcal{N}(D) + \dim \mathcal{R}(D) = \dim V = n. \tag{20}$$

Since $\left\{f_j : j \in \mathfrak{P}^c\right\}$ is contained in $\mathcal{N}(D)$ and $\left\{f_j : j \in \mathfrak{P}\right\}$ in $\mathcal{R}(D)$, it follows from Theorem 2.39 of Axler (2015) that $\left\{f_j : j \in \mathfrak{P}\right\}$ is a basis of $\mathcal{R}(D)$, and (13) follows from (18). Note that (14) follows from the observation that both $DD^{-1/2}y$ and $D^{-1/2}Dy$ equal $\sum_{j\in\mathfrak{P}}\langle y, f_j\rangle\lambda_j^{1/2}f_j = D^{1/2}y$. □

3. Results

We first present a coordinate-free definition of the multivariate Normal distribution and verify that some well-known basic facts about the multivariate Normal distribution seamlessly carry over to our coordinate-free setup. Theorem 2 shows that if Y is $\mathbb{R}^n$-valued multivariate Normal, then $(\mathcal{S}Y, \mathcal{T}Y)$ is $\mathbb{R}^m \times \mathbb{R}^p$-valued multivariate Normal, where $\mathcal{S} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathcal{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. Corollary 1, which is used in our proof of Theorem 3, then formulates a necessary and sufficient condition for the independence of $\mathcal{S}Y$ and $\mathcal{T}Y$ in terms of $\mathcal{S}$, $\mathcal{T}$, and the covariance operator of Y.

We also present alternative derivations of the independence of the sample mean and the sample variance of a random sample from a Normal distribution [Remark 2], the "partialled out" formula for population regression in the Normal model [Corollary 2], and the sufficiency of the sample mean in the Normal location model [Remark 4]. We simplify the expressions for the conditional mean and covariance obtained in Theorem 4 in Remark 3, rendering the verification of the iterated expectation and the analysis of variance formulae immediate. We outline in Remark 6 a direction in which our method can possibly be extended.

The following notational conventions and their consequences are used throughout the rest of the paper. The equality of two random variables, unless otherwise mentioned, implies equality almost surely. For any Polish space $\mathfrak{X}$, let $\mathcal{B}(\mathfrak{X})$ denote the Borel σ-algebra of $\mathfrak{X}$. Let h be a map from an arbitrary set $\mathfrak{Y}$ into a measurable space

$(\mathfrak{Z}, \mathcal{Z})$. For an arbitrary subset B of $\mathfrak{Z}$, let $h^{-1}(B)$ denote $\{y \in \mathfrak{Y} : h(y) \in B\}$, whereas for an arbitrary subset A of $\mathfrak{Y}$, let $h(A)$ denote $\{h(y) : y \in A\}$. Note that $h^{-1}(h(A)) = A$ for every subset A of $\mathfrak{Y}$. Let $\sigma(h)$ denote the smallest σ-algebra of subsets of $\mathfrak{Y}$ that makes h measurable. Since $\left\{h^{-1}(B) : B \in \mathcal{Z}\right\}$ is a σ-algebra of subsets of $\mathfrak{Y}$ (Dudley, 1989, page 98), we obtain $\sigma(h) = \left\{h^{-1}(B) : B \in \mathcal{Z}\right\}$.

Let $Z_1, \cdots, Z_n$ be iid standard Normal random variables. The distribution of

$$Z = \sum_{k=1}^{n} e_k Z_k \tag{21}$$

is defined to be the standard multivariate Normal distribution $\mathfrak{N}_n(0, I)$, where $I \in \mathcal{L}(\mathbb{R}^n)$ is the identity operator.

Lemma 3. *The characteristic function $\Psi_{0,I}$ of the $\mathfrak{N}_n(0, I)$ distribution is given by*

$$\Psi_{0,I}(t) = \mathbb{E}(\exp(i\langle t, Z\rangle)) = \exp\left(-\|t\|^2/2\right). \tag{22}$$

Proof of Lemma 3. Follows from Proposition 9.4.2(a) of Dudley (1989). □

Lemma 4. *Given $Z \sim \mathfrak{N}_n(0, I)$, $\mu \in \mathbb{R}^n$, and $T \in \mathcal{L}(\mathbb{R}^n)$, let*

$$Y = \mu + TZ. \tag{23}$$

Then, for any $s, t \in \mathbb{R}^n$,

$$\begin{aligned}
&\mathbb{E}(\exp(i\langle t, Y\rangle)) = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}\left\langle t, TT^* t\right\rangle\right) \\
&\mathbb{E}(\langle t, Y\rangle) = \langle t, \mu\rangle \\
&\mathrm{Cov}(\langle t, Y\rangle, \langle s, Y\rangle) = \left\langle t, TT^* s\right\rangle.
\end{aligned} \tag{24}$$

Proof of Lemma 4. Straightforward algebra using the bilinearity of the inner product and the definitions of T^*, Z in (21), and Y in (23), along with the fact that $Z_1, \cdots, Z_n$ are iid standard Normal random variables, proves the lemma. □

Definition 1. The distribution of Y in (23) is defined to be the multivariate Normal distribution with mean vector μ and covariance operator TT^*. Recall that a distribution on $\mathbb{R}^n$ is uniquely determined by its characteristic function (Dudley, 1989, Theorem 9.5.1); since the characteristic function of Y is determined by its mean μ and covariance TT^*, the multivariate Normal distribution $\mathfrak{N}_n(\mu, D)$ with mean μ and covariance D (for any $D \in \mathcal{L}(\mathbb{R}^n)$ positive) is uniquely defined in terms of the characteristic function

$$\Psi_{\mu,D}(t) = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}\langle t, Dt\rangle\right). \tag{25}$$

Lemma 5. *Given $Y \sim \mathfrak{N}_n(\mu, D)$ and $S \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$,*

$$SY \sim \mathfrak{N}_m\left(S\mu, SDS^*\right). \tag{26}$$

Proof of Lemma 5. The proof is a straightforward consequence of Definition 1. □

Recall that given $(s_1, t_1), (s_2, t_2) \in \mathbb{R}^m \times \mathbb{R}^p = \mathbb{R}^{m+p}$, the inner product on $\mathbb{R}^m \times \mathbb{R}^p$ is given by $\langle (s_1, t_1), (s_2, t_2) \rangle = \langle s_1, s_2 \rangle + \langle t_1, t_2 \rangle$, where the inner products on different Euclidian spaces are all denoted by $\langle \cdot, \cdot \rangle$.

Theorem 2. *Given $Y \sim \mathfrak{N}_n(\mu, D)$, $\mathcal{S} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and $\mathcal{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$, the random vector $(\mathcal{S}Y, \mathcal{T}Y) \sim \mathfrak{N}_{m+p}$ with mean $= (\mathcal{S}\mu, \mathcal{T}\mu) \in \mathbb{R}^m \times \mathbb{R}^p$ and covariance operator $\mathcal{K} \in \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^p)$ given by*

$$\mathcal{K}(s,t) = \left(\mathcal{S}D\mathcal{S}^* s + \mathcal{S}D\mathcal{T}^* t, \mathcal{T}D\mathcal{S}^* s + \mathcal{T}D\mathcal{T}^* t\right). \tag{27}$$

Proof of Theorem 2. We first verify that $\mathcal{K} : \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}^m \times \mathbb{R}^p$ is a positive operator. The verification of $\mathcal{K} \in \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^p)$ being linear and self-adjoint is routine. Since

$$\begin{aligned} &\left\langle \mathcal{K}(s,t), (s,t) \right\rangle \\ &= \left\langle \mathcal{S}D\mathcal{S}^* s, s \right\rangle + \left\langle \mathcal{S}D\mathcal{T}^* t, s \right\rangle + \left\langle \mathcal{T}D\mathcal{S}^* s, t \right\rangle + \left\langle \mathcal{T}D\mathcal{T}^* t, t \right\rangle \\ &= \left\| D^{1/2}\mathcal{S}^* s \right\|^2 + 2\left\langle D^{1/2}\mathcal{T}^* t, D^{1/2}\mathcal{S}^* s \right\rangle + \left\| D^{1/2}\mathcal{T}^* t \right\|^2 \\ &= \left\| D^{1/2}\mathcal{S}^* s + D^{1/2}\mathcal{T}^* t \right\|^2, \end{aligned} \tag{28}$$

the non-negativity of $\left\langle \mathcal{K}(s,t), (s,t) \right\rangle$ for every $(s,t) \in \mathbb{R}^m \times \mathbb{R}^p$ follows. By the definition of the inner product in $\mathbb{R}^m \times \mathbb{R}^p$ and (25), the characteristic function Ψ of the random vector $(\mathcal{S}Y, \mathcal{T}Y)$ taking values in $\mathbb{R}^m \times \mathbb{R}^p$ is given by

$$\Psi(s,t) = \exp\left(i\left\langle \mathcal{S}^* s + \mathcal{T}^* t, \mu \right\rangle - \frac{1}{2}\left\langle \mathcal{S}^* s + \mathcal{T}^* t, D\left(\mathcal{S}^* s + \mathcal{T}^* t\right) \right\rangle \right). \tag{29}$$

Since $\left\langle \mathcal{S}^* s + \mathcal{T}^* t, \mu \right\rangle = \left\langle (s,t), (\mathcal{S}\mu, \mathcal{T}\mu) \right\rangle$ and

$$\begin{aligned} &\left\langle \mathcal{S}^* s + \mathcal{T}^* t, D\left(\mathcal{S}^* s + \mathcal{T}^* t\right) \right\rangle \\ &= \left\langle s, \mathcal{S}D\mathcal{S}^* s + \mathcal{S}D\mathcal{T}^* t \right\rangle + \left\langle t, \mathcal{T}D\mathcal{S}^* s + \mathcal{T}D\mathcal{T}^* t \right\rangle \\ &= \left\langle (s,t), \left(\mathcal{S}D\mathcal{S}^* s + \mathcal{S}D\mathcal{T}^* t, \mathcal{T}D\mathcal{S}^* s + \mathcal{T}D\mathcal{T}^* t\right) \right\rangle, \end{aligned} \tag{30}$$

the proof follows. □

Corollary 1. *Given $Y \sim \mathfrak{N}_n(\mu, D)$, $\mathcal{S} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and $\mathcal{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$, $\mathcal{S}Y$ and $\mathcal{T}Y$ are independent if and only if $\mathcal{S}D\mathcal{T}^* \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^m)$, equivalently $\mathcal{T}D\mathcal{S}^* \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$, is the zero operator.*

Proof of Corollary 1. From (29),

$$\Psi(s,t) = \exp\left(i\left\langle s, \mathcal{S}\mu \right\rangle\right) \exp\left(i\left\langle t, \mathcal{T}\mu \right\rangle\right) \exp\left(-\frac{1}{2}\left\langle \mathcal{S}^* s + \mathcal{T}^* t, D\left(\mathcal{S}^* s + \mathcal{T}^* t\right) \right\rangle\right). \tag{31}$$

Now by (30),

$$\left\langle S^*s + \mathcal{T}^*t, D\left(S^*s + \mathcal{T}^*t\right)\right\rangle = \left\langle s, SDS^*s\right\rangle + 2\left\langle s, SD\mathcal{T}^*t\right\rangle + \left\langle t, \mathcal{T}D\mathcal{T}^*t\right\rangle. \quad (32)$$

Thus, by (25) and Lemma 5,

$$\Psi(s,t) = \mathbb{E}\left(\exp\left(i\left\langle s, SY\right\rangle\right)\right)\mathbb{E}\left(\exp\left(i\left\langle t, \mathcal{T}Y\right\rangle\right)\right)\exp\left(-\left\langle s, SD\mathcal{T}^*t\right\rangle\right). \quad (33)$$

That is, the characteristic function of $(SY, \mathcal{T}Y)$ is the product of two factors; one factor is the characteristic function of the product measure of the distributions induced by SY on $\mathcal{B}(\mathbb{R}^m)$ and $\mathcal{T}Y$ on $\mathcal{B}(\mathbb{R}^p)$, whereas the other factor is $\exp\left(-\left\langle s, SD\mathcal{T}^*t\right\rangle\right)$. Therefore, SY and $\mathcal{T}Y$ are independent if and only if $\exp\left(-\left\langle s, SD\mathcal{T}^*t\right\rangle\right) = 1$ for every $s \in \mathbb{R}^m$ and $t \in \mathbb{R}^p$, equivalently, $\left\langle s, SD\mathcal{T}^*t\right\rangle = 0$ for every s and t. Since $\mathcal{T}DS^* = (SD\mathcal{T}^*)^*$ by Theorems 7.6(e) and 7.6(c) of Axler (2015), the proof follows. □

Remark 2. For $X_1, \cdots, X_n$ iid Normal random variables with mean θ and variance σ^2,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \quad (34)$$

are independent (Casella and Berger, 2002, Theorem 5.3.1(a)). Most textbooks prove this result by working with the (joint) density of the sample and using the Jacobian formula for finding the density of the transformation that maps the sample to the sample mean and the sample variance. Some textbooks use Basu's Theorem (Basu, 1955) on an ancillary statistic (the sample variance) being independent of a complete sufficient statistic (the sample mean) to prove this result. We are going to show that this result is a straightforward consequence of Corollary 1.

Let J be the sum of the standard orthonormal basis vectors, and $\{J\}$ the span of J. Since

$$\Pi_{\{J\}}x = \|J\|^{-2}\langle x, J\rangle J, \quad (35)$$

we have

$$\bar{X} = \|J\|^{-2}\left\langle \Pi_{\{J\}}X, J\right\rangle \text{ and } S^2 = \left(\|J\|^2 - 1\right)^{-1}\left\|\left(I - \Pi_{\{J\}}\right)X\right\|^2, \quad (36)$$

where $X = \sum_{k=1}^{n} X_k e_k \sim \mathfrak{N}_n\left(\theta J, \sigma^2 I\right)$. By Corollary 1, using the fact that an orthogonal projection operator is self-adjoint, we obtain $\Pi_{\{J\}}Y$ and $\left(I - \Pi_{\{J\}}\right)Y$ are independent if and only if $\left(I - \Pi_{\{J\}}\right)D\Pi_{\{J\}}$ is the zero operator, where $Y \sim \mathfrak{N}_n(\mu, D)$. Clearly, $\left(I - \Pi_{\{J\}}\right)D\Pi_{\{J\}}x = 0$ for all $x \in \mathbb{R}^n$ if and only if $\left(I - \Pi_{\{J\}}\right)DJ = 0$, equivalently, J is an eigenvector of D corresponding to the

eigenvalue $\|J\|^{-2}\langle DJ, J\rangle$. Since J is an eigenvector of $\sigma^2 I$, the independence of $\bar{X}$ and S^2 follows from that of $\Pi_{\{J\}}X$ and $\left(I - \Pi_{\{J\}}\right)X$.

It is interesting to note that the class of positive operators with J as an eigenvector does not reduce to the singleton set $\{I\}$. For $n = 2$, define the positive $D \in \mathcal{L}\left(\mathbb{R}^2\right)$ by $D\left(u_1, u_2\right) = \left(2u_1 + u_2, u_1 + 2u_2\right)$; clearly, $J = (1, 1)$ is an eigenvector for D. Thus, for the sample mean and the sample variance to be independent, it is not necessary for the sample to be iid. As long as the joint distribution of the sample is multivariate Normal such that J is an eigenvector of the covariance operator, the independence of the sample mean and the sample variance holds. However, for Normal random variables that are dependent, whether the joint distribution is multivariate Normal becomes a modeling question, as the joint distribution of even pairwise uncorrelated Normal random variables may not be multivariate Normal.

We are now ready to present the two main results.

Theorem 3. *Given $Y \sim \mathfrak{N}_n(\mu, D)$ and $T \in \mathcal{L}(\mathbb{R}^n)$, define*

$$S = TD^{1/2}; \tag{37}$$

then

$$\widehat{T}Y = D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}Y \text{ is independent of } TY \tag{38}$$

and

$$Y - D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}Y \text{ is an affine function of } TY. \tag{39}$$

Proof of Theorem 3. For any $x \in \mathbb{R}^n$ and $F \in \mathcal{L}(\mathbb{R}^n)$, we obtain from (14),

$$TD\left(D^{1/2}\Pi_{\mathcal{N}(F)}D^{-1/2}\right)^* x = TDD^{-1/2}\Pi_{\mathcal{N}(F)}D^{1/2}x = TD^{1/2}\Pi_{\mathcal{N}(F)}D^{1/2}x, \tag{40}$$

implying, by (37), that $TD\left(D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}\right)^*$ is the zero operator, whence (38) follows from Corollary 1.

To prove (39) we first observe that, by (19) and (13) in that order,

$$Y = \Pi_{\mathcal{R}(D)}Y + \Pi_{\mathcal{N}(D)}Y = D^{1/2}D^{-1/2}Y + \Pi_{\mathcal{N}(D)}Y. \tag{41}$$

Now we are going to show that

$$\Pi_{\mathcal{N}(D)}Y = \Pi_{\mathcal{N}(D)}\mu, \tag{42}$$

implying, by (41), that

$$Y = D^{1/2}D^{-1/2}Y + \Pi_{\mathcal{N}(D)}\mu. \tag{43}$$

Let f_j, λ_j, $\mathfrak{P}$, and $\mathfrak{P}^c$ be as in the proof of Lemma 2. Recall that $\left\{f_j : j \in \mathfrak{P}^c\right\}$ is an orthonormal basis for $\mathcal{N}(D)$ and $\left\{f_j : j \in \mathfrak{P}\right\}$ is an orthonormal basis for $\mathcal{R}(D)$. For $j \in \mathfrak{P}^c$ and $u \in \mathbb{R}$, by (25),

$$\mathbb{E}\left(\exp\left(iu\left\langle Y, f_j\right\rangle\right)\right) = \exp\left(i\left\langle uf_j, \mu\right\rangle - \frac{1}{2}\left\langle uf_j, Duf_j\right\rangle\right) = \exp\left(iu\left\langle \mu, f_j\right\rangle\right), \tag{44}$$

implying $\left\langle Y, f_j\right\rangle = \left\langle \mu, f_j\right\rangle$ for all $j \in \mathfrak{P}^c$, that is, (42).

By Theorems 6.47 and 7.7(c) of Axler (2015),

$$I = \Pi_{\mathcal{R}(S^*)} + \Pi_{\mathcal{N}(S)}, \tag{45}$$

implying, by (43),

$$Y - D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}Y = D^{1/2}\Pi_{\mathcal{R}(S^*)}D^{-1/2}Y + \Pi_{\mathcal{N}(D)}\mu. \tag{46}$$

By Theorem 1, there exists an invertible $U \in \mathcal{L}(\mathbb{R}^n)$ such that

$$\Pi_{\mathcal{R}(S^*)} = US; \tag{47}$$

applying (47), (37), and (43) in that order, we obtain that RHS(46) equals

$$D^{1/2}USD^{-1/2}Y + \Pi_{\mathcal{N}(D)}\mu = D^{1/2}UTY + \left(I - D^{1/2}UT\right)\Pi_{\mathcal{N}(D)}\mu, \tag{48}$$

thereby establishing (39). □

Theorem 4. *Given $Y \sim \mathfrak{N}_n(\mu, D)$ and $\mathcal{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, the conditional distribution of Y given $\mathcal{T}Y$ is multivariate Normal on $\mathcal{B}(\mathbb{R}^n)$.*

The following lemma on measurability is used in the proof of Theorem 4.

Lemma 6. *Let $\mathfrak{X}$ and $\mathfrak{Y}$ be Polish spaces. If $h : \mathfrak{X} \mapsto \mathfrak{Y}$ is Borel measurable and injective, then $\sigma(h) = \mathcal{B}(\mathfrak{X})$.*

Proof of Lemma 6. The inclusion $\sigma(h) \subseteq \mathcal{B}(\mathfrak{X})$ follows from the pertinent definitions. To establish the reverse inclusion, fix $M \in \mathcal{B}(\mathfrak{X})$ arbitrarily. Since $M = h^{-1}(h(M))$ and $h(M) \in \mathcal{B}(\mathfrak{Y})$ by Theorem I.3.9 of Parthasarathy (1967), $M \in \sigma(h)$. □

We now present the proof of Theorem 4.

Proof of Theorem 4. We first construct $T \in \mathcal{L}(\mathbb{R}^n)$ such that $\sigma(\mathcal{T}Y) = \sigma(TY)$, and then use Theorem 3 to find the conditional distribution of Y given TY.

Let $\mathfrak{T} \in \mathcal{L}(\mathbb{R}^n, \mathcal{R}(\mathcal{T}))$ be defined by $\mathfrak{T}x = \mathcal{T}x$ for every $x \in \mathbb{R}^n$. Clearly, $\mathfrak{T}$ is surjective, hence injective and invertible (Axler, 2015, Theorem 3.69). By Theorem 7.7 of Axler (2015), $\mathfrak{T}^* \in \mathcal{L}(\mathcal{R}(\mathcal{T}), \mathbb{R}^n)$ is invertible as well. Let $T \in \mathcal{L}(\mathbb{R}^n)$ be defined by $Tx = \mathfrak{T}^*\mathfrak{T}x$; note that T, being the composition of two invertible transformations, is invertible.

Let $(\Omega, \mathcal{F}, P)$ denote the probability space underlying Y. Since $\mathcal{T}$ is continuous, hence measurable, $\mathcal{T}Y : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ is measurable, implying $(\mathcal{T}Y)^{-1}(E) \in \mathcal{F}$ for every $E \in \mathcal{B}(\mathbb{R}^m)$ and $\sigma(\mathcal{T}Y) = \left\{(\mathcal{T}Y)^{-1}(E) : E \in \mathcal{B}(\mathbb{R}^m)\right\}$. Now note that $(\mathcal{T}Y)^{-1}(E) = (\mathfrak{T}Y)^{-1}(E \cap \mathcal{R}(\mathcal{T}))$, and $\mathcal{R}(\mathcal{T})$, by virtue of being closed, is an element of $\mathcal{B}(\mathbb{R}^m)$, whence $\mathcal{B}(\mathcal{R}(\mathcal{T}))$ is the trace of $\mathcal{B}(\mathbb{R}^m)$ on $\mathcal{R}(\mathcal{T})$; that is, $\mathcal{B}(\mathcal{R}(\mathcal{T})) = \{E \cap \mathcal{R}(\mathcal{T}) : E \in \mathcal{B}(\mathbb{R}^m)\}$, implying

$$\begin{aligned}\sigma(\mathcal{T}Y) = \left\{(\mathcal{T}Y)^{-1}(E) : E \in \mathcal{B}(\mathbb{R}^m)\right\} &= \left\{(\mathfrak{T}Y)^{-1}(E \cap \mathcal{R}(\mathcal{T})) : E \in \mathcal{B}(\mathbb{R}^m)\right\} \\ &= \left\{(\mathfrak{T}Y)^{-1}(F) : F \in \mathcal{B}(\mathcal{R}(\mathcal{T}))\right\} \\ &= \sigma(\mathfrak{T}Y).\end{aligned} \tag{49}$$

Since $\mathfrak{T}$ and T are injective (and measurable), $\sigma(\mathfrak{T}) = \mathcal{B}(\mathbb{R}^n) = \sigma(T)$ by Lemma 6; that is, $\left\{\mathfrak{T}^{-1}(G) : G \in \mathcal{B}(\mathcal{R}(\mathcal{T}))\right\} = \mathcal{B}(\mathbb{R}^n) = \left\{T^{-1}(F) : F \in \mathcal{B}(\mathbb{R}^n)\right\}$. Since, for any $G \in \mathcal{B}(\mathcal{R}(\mathcal{T}))$ and $F \in \mathcal{B}(\mathbb{R}^n)$,

$$(\mathfrak{T}Y)^{-1}(G) = Y^{-1}\left(\mathfrak{T}^{-1}(G)\right) \text{ and } (TY)^{-1}(F) = Y^{-1}\left(T^{-1}(F)\right), \tag{50}$$

we obtain

$$\begin{aligned}\sigma(\mathfrak{T}Y) &= \left\{(\mathfrak{T}Y)^{-1}(G) : G \in \mathcal{B}(\mathcal{R}(\mathcal{T}))\right\} \\ &= \left\{(TY)^{-1}(F) : F \in \mathcal{B}(\mathbb{R}^n)\right\} = \sigma(TY).\end{aligned} \tag{51}$$

That establishes the equality of $\sigma(\mathcal{T}Y)$ and $\sigma(TY)$.

By the pull-out property of conditional expectation (Kallenberg, 2002, page 105), (38), and (39),

$$\begin{aligned}&\mathbb{E}\left(\exp\left(i\langle t, Y\rangle\right)\middle|TY\right) \\ &= \exp\left(i\langle t, Y - D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}Y\rangle\right)\mathbb{E}\left(\exp\left(i\langle t, D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}Y\rangle\right)\right),\end{aligned} \tag{52}$$

where S is as in (37). By Lemma 5 and (25),

$$\begin{aligned}&\mathbb{E}\left(\exp\left(i\langle t, D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}Y\rangle\right)\right) \\ &= \exp\left(i\langle t, D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}\mu\rangle - \frac{1}{2}\langle t, Gt\rangle\right),\end{aligned} \tag{53}$$

where the operator $G = D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}DD^{-1/2}\Pi_{\mathcal{N}(S)}D^{1/2}$ equals, by (14) and (13), $D^{1/2}\Pi_{\mathcal{N}(S)}\Pi_{\mathcal{R}(D)}\Pi_{\mathcal{N}(S)}D^{1/2}$. Therefore, the conditional distribution of Y given TY, hence the conditional distribution given $\mathcal{T}Y$, is Normal with mean $\upsilon(Y) = Y - D^{1/2}\Pi_{\mathcal{N}(S)}D^{-1/2}(Y - \mu)$ and covariance G. □

Remark 3. The expressions for the mean υ and the covariance G of the conditional distribution of Y given $\mathcal{T}Y$ can be considerably simplified, rendering the verification of the iterated expectation formula and the analysis of variance formula immediate.

Since $Y-\mu=\Pi_{\mathcal{R}(D)}(Y-\mu)$ by (19) and (42), applying (45) and (13) in that order,

$$\upsilon(Y)=\mu+D^{1/2}\Pi_{\mathcal{R}(S^*)}D^{-1/2}(Y-\mu). \tag{54}$$

To simplify the expression of the conditional covariance operator G (which does not depend on Y), first note that, by (13) and (14),

$$D^{1/2}\Pi_{\mathcal{R}(D)}=D^{1/2}D^{1/2}D^{-1/2}=DD^{-1/2}=D^{1/2}. \tag{55}$$

By (47), (37), (55), (37), and (47), in that order,

$$\Pi_{\mathcal{R}(S^*)}\Pi_{\mathcal{R}(D)}=USU\Pi_{\mathcal{R}(D)}=UTD^{1/2}\Pi_{\mathcal{R}(D)}=UTD^{1/2}=US=\Pi_{\mathcal{R}(S^*)}. \tag{56}$$

Consequently, G equals

$$\begin{aligned}
&D^{1/2}\Pi_{\mathcal{R}(D)}\Pi_{\mathcal{N}(S)}D^{1/2}-D^{1/2}\Pi_{\mathcal{R}(S^*)}\Pi_{\mathcal{R}(D)}\Pi_{\mathcal{N}(S)}D^{1/2} && \text{by (45)}\\
&=D^{1/2}\Pi_{\mathcal{N}(S)}D^{1/2}-D^{1/2}\Pi_{\mathcal{R}(S^*)}\Pi_{\mathcal{R}(D)}\Pi_{\mathcal{N}(S)}D^{1/2} && \text{by (55)}\\
&=D^{1/2}\Pi_{\mathcal{N}(S)}D^{1/2}-D^{1/2}\Pi_{\mathcal{R}(S^*)}\Pi_{\mathcal{N}(S)}D^{1/2} && \text{by (56)}\\
&=D^{1/2}\Pi_{\mathcal{N}(S)}D^{1/2} && \text{by (45)}.
\end{aligned} \tag{57}$$

Recall that if V and W are random vectors in $\mathbb{R}^n$ such that V is an affine function of W, i.e., $V=a+RW$, where $a\in\mathbb{R}^n$ and $R\in\mathcal{L}(\mathbb{R}^n)$, then

$$\mathbb{E}(V)=a+R(\mathbb{E}(W)) \text{ and } \mathrm{Cov}(V)=R\mathrm{Cov}(W)R^*, \tag{58}$$

implying $\mathbb{E}(\mathbb{E}(Y|TY))=\mu$ by (54), verifying the iterated expectation formula.

Defining the expected value of an operator-valued random element Υ as the operator $\mathbb{E}(\Upsilon)$ such that $\mathbb{E}(\Upsilon)(s)=\mathbb{E}(\Upsilon(s))$ for every $s\in\mathbb{R}^n$, provided the expected value on the right hand side is well defined for every s, we conclude from (57) that

$$\mathbb{E}(\mathrm{Cov}(Y|TY))=D^{1/2}\Pi_{\mathcal{N}(S)}D^{1/2}. \tag{59}$$

By (54) and Lemma 5,

$$\mathrm{Cov}(\mathbb{E}(Y|TY))=D^{1/2}\Pi_{\mathcal{R}(S^*)}D^{-1/2}DD^{-1/2}\Pi_{\mathcal{R}(S^*)}D^{1/2}=D^{1/2}\Pi_{\mathcal{R}(S^*)}D^{1/2}, \tag{60}$$

where the second equality follows by (14), (13), (56), and idempotence of $\Pi_{\mathcal{R}(S^*)}$, in that order. Clearly, (45) implies $\mathbb{E}(\mathrm{Cov}(Y|TY))+\mathrm{Cov}(\mathbb{E}(Y|TY))=\mathrm{Cov}(Y)$, thereby verifying the analysis of variance formula.

An immediate corollary of Theorem 4 is the "partialled out" formula (Wooldridge, 2013, page 78) for population regression in the Normal model.

Corollary 2. *If* (X, W, Z) *is multivariate Normal such that* $Z = \left(Z_1, \cdots, Z_{n-2}\right) \in \mathbb{R}^{n-2}$ *and* $X, W \in \mathbb{R}$*, then*

$$\mathbb{E}(W|X,Z) - \mathbb{E}(W|Z) = \frac{\mathrm{Cov}(X,W|Z)}{\mathrm{Var}(X|Z)} I_{[\mathrm{Var}(X|Z)>0]}(X - \mathbb{E}(X|Z)), \tag{61}$$

where the indicator function of a set A is denoted by I_A*, and* $I_{[\mathrm{Var}(X|Z)>0]}/\mathrm{Var}(X|Z)$ *denotes the measurable function of* Z *that equals* 0 *if* $\mathrm{Var}(X|Z) = 0$ *and* $1/\mathrm{Var}(X|Z)$ *if* $\mathrm{Var}(X|Z) > 0$*.*

The following lemma on projection operators is used in the proof of Corollary 2.

Lemma 7. *For a proper and closed subspace* V *of a Hilbert space* H *and* $x \in H \backslash V$*, let* V_x *denote the closure of the subspace* $\{v + cx : v \in V, c \in \mathbb{R}\}$*. Then, for any* $y \in H$*,*

$$\Pi_{V_x} y - \Pi_V y = \left\|\Pi_{V^\perp} x\right\|^{-2} \left\langle \Pi_{V^\perp} y, \Pi_{V^\perp} x \right\rangle \Pi_{V^\perp} x. \tag{62}$$

Proof of Lemma 7. By the definition of V_x there exists a sequence $\left\{v_n : n \geq 1\right\} \subset V$ and a sequence $\left\{c_n : n \geq 1\right\} \subset \mathbb{R}$ such that

$$\Pi_{V_x} y = \lim_{n\to\infty} \left(v_n + c_n x\right). \tag{63}$$

Since $V \subset V_x$, $\Pi_V y = \Pi_V \Pi_{V_x} y$; since Π_V is continuous, by (63),

$$\Pi_V y = \lim_{n\to\infty} \left(v_n + c_n \Pi_V x\right). \tag{64}$$

Since $x - \Pi_V x = \Pi_{V^\perp} x$, subtracting (64) from (63),

$$\mathrm{LHS}(62) = \lim_{n\to\infty} \left(c_n \Pi_{V^\perp} x\right); \tag{65}$$

since $x \notin V$, that is, $\Pi_{V^\perp} x \neq 0$, and $\mathbb{R}$ is complete, $\lim_{n\to\infty} c_n$ exists, implying

$$\mathrm{LHS}(62) = \left(\lim_{n\to\infty} c_n\right) \Pi_{V^\perp} x. \tag{66}$$

Since $\mathrm{LHS}(62) = \Pi_{V^\perp} y - \Pi_{(V_x)^\perp} y$, taking the inner product of both sides of (66) with $\Pi_{V^\perp} x$ and using the linearity and homogeneity of inner product, we obtain

$$\left\langle \Pi_{V^\perp} y, \Pi_{V^\perp} x \right\rangle - \left\langle \Pi_{(V_x)^\perp} y, \Pi_{V^\perp} x \right\rangle = \left(\lim_{n\to\infty} c_n\right) \left\|\Pi_{V^\perp} x\right\|^2. \tag{67}$$

Since $x \in V_x$ and $\Pi_V x \in V \subset V_x$, $\Pi_{V^\perp} x \in V_x$, implying $\left\langle \Pi_{(V_x)^\perp} y, \Pi_{V^\perp} x \right\rangle = 0$; since $\left\|\Pi_{V^\perp} x\right\|^2 \neq 0$, the lemma follows from (67) and (66). □

We now present the proof of Corollary 2.

Proof of Corollary 2. Let $\mathcal{P}_{1,3} \in \mathcal{L}\left(\mathbb{R}^n, \mathbb{R}^{n-1}\right)$ and $\mathcal{P}_3 \in \mathcal{L}\left(\mathbb{R}^n, \mathbb{R}^{n-2}\right)$ be defined by $\mathcal{P}_{1,3} y = (x, z)$ and $\mathcal{P}_3 y = z$, where $y = (x, w, z)$. From the proof of Theorem 4, using (54), we obtain

$$\text{LHS}(61) = \left\langle D^{1/2}\left(\Pi_{\mathcal{R}\left(S_{1,3}^*\right)} - \Pi_{\mathcal{R}\left(S_3^*\right)}\right) D^{-1/2}(Y-\mu), e_2\right\rangle, \tag{68}$$

where $Y = (X, W, Z)$, $\mu \in \mathbb{R}^n$ is the mean of Y and $D \in \mathcal{L}(\mathbb{R}^n)$ is the covariance of Y, $S_{1,3} = P_{1,3}D^{1/2}$ with $P_{1,3} \in \mathcal{L}(\mathbb{R}^n)$ defined by $P_{1,3}y = (x, 0, z)$, and $S_3 = P_3 D^{1/2}$ with $P_3 \in \mathcal{L}(\mathbb{R}^n)$ defined by $P_3 y = (0, 0, z)$.

To show RHS(61) = RHS(68), we first observe, using (57) and (54),

$$\begin{aligned}
\text{Cov}(X, W|Z) &= \left\langle D^{1/2}\Pi_{\mathcal{N}(S_3)}D^{1/2}e_1, e_2\right\rangle \\
\text{Var}(X|Z) &= \left\langle D^{1/2}\Pi_{\mathcal{N}(S_3)}D^{1/2}e_1, e_1\right\rangle = \left\|\Pi_{\mathcal{N}(S_3)}D^{1/2}e_1\right\|^2 \\
X - \mathbb{E}(X|Z) &= \left\langle \left(I - D^{1/2}\Pi_{\mathcal{R}\left(S_3^*\right)}D^{-1/2}\right)(Y-\mu), e_1\right\rangle.
\end{aligned} \tag{69}$$

Since $P_{1,3}e_j = e_j$ if $j \neq 2$, $P_{1,3}e_2 = 0$, $P_3 e_j = e_j$ if $j \geq 3$, and $P_3 e_j = 0$ if $j = 1, 2$, we obtain, using that $P_{1,3}$, P_3, and $D^{1/2}$ are all self-adjoint,

$$\begin{aligned}
\mathcal{R}\left(S_{1,3}^*\right) &= \text{span}\left\{D^{1/2}e_1, D^{1/2}e_3, \cdots, D^{1/2}e_n\right\} \\
\mathcal{R}\left(S_3^*\right) &= \text{span}\left\{D^{1/2}e_3, \cdots, D^{1/2}e_n\right\}.
\end{aligned} \tag{70}$$

If $D^{1/2}e_1 \in \text{span}\left\{D^{1/2}e_3, \cdots, D^{1/2}e_n\right\} = \mathcal{R}\left(S_3^*\right)$, then $\mathcal{R}\left(S_{1,3}^*\right) = \mathcal{R}\left(S_3^*\right)$, implying RHS(68) = 0; in that case $\text{Var}(X|Z)$, a constant function of Z, is also equal to 0 by the second row of (69), thereby vacuously satisfying (61).

If $D^{1/2}e_1 \notin \text{span}\left\{D^{1/2}e_3, \cdots, D^{1/2}e_n\right\}$, that is, $\text{Var}(X|Z) > 0$, by Lemma 7,

$$\begin{aligned}
&\left(\Pi_{\mathcal{R}\left(S_{1,3}^*\right)} - \Pi_{\mathcal{R}\left(S_3^*\right)}\right) D^{-1/2}(Y-\mu) \\
&= \left\|\Pi_{\mathcal{N}(S_3)}D^{1/2}e_1\right\|^{-2} \left\langle \Pi_{\mathcal{N}(S_3)}D^{-1/2}(Y-\mu), \Pi_{\mathcal{N}(S_3)}D^{1/2}e_1\right\rangle \Pi_{\mathcal{N}(S_3)}D^{1/2}e_1.
\end{aligned} \tag{71}$$

Therefore, RHS(68) equals

$$\frac{\left\langle \Pi_{\mathcal{N}(S_3)}D^{-1/2}(Y-\mu), \Pi_{\mathcal{N}(S_3)}D^{1/2}e_1\right\rangle}{\left\|\Pi_{\mathcal{N}(S_3)}D^{1/2}e_1\right\|^2} \left\langle D^{1/2}\Pi_{\mathcal{N}(S_3)}D^{1/2}e_1, e_2\right\rangle, \tag{72}$$

which, by the first two rows in (69), equals, since $\Pi_{\mathcal{N}(S_3)}$ is self-adjoint and idempotent,

$$\frac{\text{Cov}(X, Y|Z)}{\text{Var}(X|Z)} \left\langle D^{1/2}\Pi_{\mathcal{N}(S_3)}D^{-1/2}(Y-\mu), e_1\right\rangle. \tag{73}$$

Since by (45), (13), and (19),

$$D^{1/2}\Pi_{\mathcal{N}(S_3)}D^{-1/2} = I - D^{1/2}\Pi_{\mathcal{R}\left(S_3^*\right)}D^{-1/2} - \Pi_{\mathcal{N}(D)} \tag{74}$$

the proof follows by the third row in (69) and (42). □

Remark 4. Given a random sample $X_1, \cdots, X_n$ from the Normal distribution with mean θ and (known) variance σ^2, $\bar{X}$ is a sufficient statistic for θ (Casella and Berger, 2002, Example 6.2.4). While the typical proof uses the powerful factorization theorem (Casella and Berger, 2002, Theorem 6.2.6), we are going to show that the conditional distribution of the sample $X = \sum_{k=1}^{n} X_k e_k \sim \mathfrak{N}_n\left(\theta J, \sigma^2 I\right)$ given $\bar{X}$, that is, $n^{-1}\Pi_{\{J\}}X$, does not depend on θ, thereby proving the sufficiency of $\bar{X}$ directly from the definition. The conditional distribution of the sample X given $n^{-1}\Pi_{\{J\}}X$, by Theorem 4, (54), and (57), is multivariate Normal with mean $\theta J + D^{1/2}\Pi_{\mathcal{R}(S^*)}D^{-1/2}(X - \theta J)$ and covariance $D^{1/2}\Pi_{\mathcal{N}(S)}D^{1/2}$, where $D = \sigma^2 I$ and $S = n^{-1}\Pi_{\{J\}}D^{1/2}$. Since $S = \sigma n^{-1}\Pi_{\{J\}}$, we successively obtain that $S^* = \sigma n^{-1}\Pi_{\{J\}}$, $\mathcal{R}(S^*) = \{J\}$, and $\Pi_{\mathcal{R}(S^*)} = \Pi_{\{J\}}$, implying that the mean equals $\Pi_{\{J\}}X$ and the covariance equals $\sigma^2\left(I - \Pi_{\{J\}}\right)$.

Remark 5. The next step in this line of research is to approximate the conditional distribution of Y given $g(Y)$, where g is a C^1 vector field on $\mathbb{R}^n$ that takes values in $\mathbb{R}^m$, by using the local linearity of g and the decomposition of Y in Theorem 3. If g is constant, then the σ-algebra generated by $g(Y)$ is the trivial σ-algebra consisting of the empty set and the entire sample space, making Y independent of $g(Y)$, so that the conditional distribution of Y given $g(Y)$ is the unconditional distribution of Y. If g is injective, $\sigma(g)$ equals $\mathcal{B}(\mathbb{R}^n)$ by Lemma 6; consequently, $\sigma(g(Y))$ equals $\sigma(Y)$, implying that the conditional distribution of Y given $g(Y)$ is the point mass at Y. Thus, the interesting problem unfolds when g is non-constant and non-injective.

Let $\xi : \mathbb{R}^n \mapsto \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ map $a \in \mathbb{R}^n$ to the total derivative of g at a. Let $J(a)$ denote the $m \times n$ matrix that represents $\xi(a)$ with respect to the standard orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. Since g is in C^1, for every $1 \leq i \leq m$ and $1 \leq k \leq n$, the function $a \mapsto J_{ik}(a)$ is continuous from $\mathbb{R}^n$ to $\mathbb{R}$. Since the operator norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is equivalent to the Euclidian norm on $\mathbb{R}^{mn}$, ξ is continuous.

Given $\epsilon > 0$, there exists a compact subset K_ϵ of $\mathbb{R}^n$ such that $P\left(Y \in K_\epsilon\right) > 1 - \epsilon$. Restricted to K_ϵ, the function ξ is uniformly continuous. Consequently, it admits a modulus of continuity ζ_ϵ which, restricted to the closed interval $\left[0, B_\epsilon\right]$, where B_ϵ is the diameter of K_ϵ, is uniformly continuous. We are working to show that, for a suitably chosen metric for weak convergence of probability measures on $\mathbb{R}^n$ and $\delta > 0$ (that depends on the given ϵ through the supremum of the modulus of continuity ζ_ϵ on $\left[0, B_\epsilon\right]$), the conditional distribution of Y given $g(Y)$ is in a δ-neighborhood of the family of multivariate Normal distributions.

Remark 6. Proposition 3.13 of Eaton (1983) holds at a much greater level of generality; see Bogachev (1998, Theorem 3.10.1). Extending our approach, in the absence of an inner product, to finding the conditional distribution of a Gaussian random element given a (non-injective, non-constant) linear transformation appears to be an interesting area of future research. In particular, if Y is the Brownian motion,

 https://doi.org/10.1016/j.heliyon.2019.e01136

i.e., the random element in $C([0,1])$ distributed according to the Wiener measure, $\nu_1, \cdots, \nu_k$ are measures on $\mathcal{B}([0,1])$, and $\mathcal{T} : C([0,1]) \mapsto \mathbb{R}^k$ is given by $\mathcal{T}f = \left(\int f d\nu_1, \cdots, \int f d\nu_k\right)$, what is the conditional distribution of Y given $\mathcal{T}Y$?

Declarations

Author contribution statement

Rajeshwari Majumdar, Suman Majumdar: Conceived and designed the analysis; Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis, 2nd ed. John Wiley & Sons, New York, NY.

Axler, Sheldon, 2015. Linear Algebra Done Right, 3rd ed. Springer, New York, NY.

Basu, D., 1955. On statistics independent of a complete sufficient statistic. Sankhya 15, 377–380.

Bogachev, Vladimir I., 1998. Gaussian Measures, 1st ed. American Mathematical Society, Providence, RI.

Casella, George, Berger, Roger L., 2002. Statistical Inference, 2nd ed. Brooks/Cole Cengage Learning, Belmont, CA.

Dudley, R.M., 1989. Real Analysis and Probability, 1st ed. Wadsworth and Brooks/Cole, Pacific Grove, CA.

 https://doi.org/10.1016/j.heliyon.2019.e01136

Eaton, Morris L., 1983. Multivariate Statistics: A Vector Space Approach, 1st ed. John Wiley & Sons, New York, NY.

Flury, Bernard, 1997. A First Course in Multivariate Statistics, 1st ed. Springer, New York, NY.

Kallenberg, Olav, 2002. Foundations of Modern Probability, 2nd ed. Springer, New York, NY.

Muirhead, Robb J., 1982. Aspects of Multivariate Statistical Theory, 1st ed. John Wiley & Sons, New York, NY.

Parthasarathy, K.R., 1967. Probability Measures on Metric Spaces. Academic Press, New York, NY.

Penrose, Roger, 1955. A generalized inverse for matrices. Proc. Camb. Philos. Soc. 51, 406–413.

Wooldridge, Jeffrey M., 2013. Introductory Econometrics: A Modern Approach. South-Western Cengage Learning, Mason, OH.

 https://doi.org/10.1016/j.heliyon.2019.e01136