Lecture 22: A Review of Linear Algebra and an Introduction to The Multivariate Normal Distribution

Relevant textbook passages:

Larsen–Marx [8]: Section 10.5.

Pitman [9]: Section 6.5.

If you just need a quick refresher, I recommend my on-line notes [4, 5]. In what follows, vectors in $\boldsymbol{R}^{\mathrm{n}}$ are usually considered to be $n \times 1$ column matrices, and $'$ denotes transposition. To save space I may write a vector a $x = (x_1, \dots, x_n)$, but you should still think of it as a column vector.

22.1 The geometry of the Euclidean inner product

Vectors x and y in $\boldsymbol{R}^{\mathrm{n}}$ are **orthogonal** if $x'y = 0$, written as $x \perp y$.

More generally, for nonzero vectors x and y in a Euclidean space,

$$x'y = x \cdot y = \|x\| \, \|y\| \cos\theta,$$

where θ is the angle between x and y.

To see this, orthogonally project y on the space spanned by x. That is, write $y = \alpha x + z$ where $z \cdot x = 0$. Thus

$$z \cdot x = (y - \alpha x) \cdot x = y \cdot x - \alpha x \cdot x = 0 \quad \Longrightarrow \quad \alpha = x \cdot y / x \cdot x.$$

Referring to Figure 22.1 we see that

$$\cos\theta = \alpha \|x\| / \|y\| = x \cdot y / \left(\|x\| \, \|y\| \right).$$

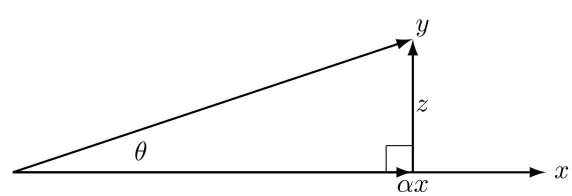

Figure 22.1. Dot product and angles: $\cos\theta = \alpha\|x\|/\|y\| = x \cdot y / \left(\|x\| \, \|y\|\right)$.

22.2 Reminder of matrix operations

Here is a quick reminder of some of the properties of matrix operations that we will use often.

- $(AB)' = B'A'$.

- $A(B+C)D = ABD + ACD$.
- Operations can proceed blockwise:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{bmatrix}$$

- For an $n \times 1$ column matrix (an n-vector) the matrix xx' is the $n \times n$ matrix whose i,j-element is $x_i x_j$.

22.3 Orthogonal matrices

22.3.1 Definition *Let A be an $n \times n$ square matrix. We say that A is an* ***orthogonal matrix*** *if its transpose is its inverse,*

$$A' = A^{-1}.$$

This may seem like an odd property to study, but the following theorem explains why it is so useful. Essentially an orthogonal matrix rotates (or reflects) the coordinate system without distorting distances or angles.

22.3.2 Proposition *For an $n \times n$ square matrix A, the following are equivalent.*

1. *A is orthogonal. That is, $A'A = I$.*
2. *A preserves norms. That is, for all x,*

$$\|Ax\| = \|x\|. \tag{1}$$

3. *A preserves inner products, that is, for every $x, y \in \boldsymbol{R}^{\mathrm{n}}$,*

$$(Ax) \cdot (Ay) = x \cdot y$$

Proof: (1) $\implies$ (2) Assume $A' = A^{-1}$. For any vector, $\|x\|^2 = x'x$, so

$$\|Ax\|^2 = (Ax)'(Ax) = x'A'Ax = x'Ix = x'x = \|x\|^2.$$

(2) $\implies$ (3) Assume A preserves norms. By Lemma 22.3.3 below,

$$\begin{aligned}(Ax) \cdot (Ay) &= \frac{\|Ax + Ay\| - \|Ax - Ay\|}{4} \\ &= \frac{\|A(x+y)\| - \|A(x-y)\|}{4} \\ &= \frac{\|x+y\| - \|x-y\|}{4} \\ &= x'y.\end{aligned}$$

(3) $\implies$ (1) Assume A preserves inner products. Pick an arbitrary x and z, and let $y = Az$. Then $y'Ax = z'A'Ax = z \cdot x$ since A preserves inner products. That is, $z'A'Ax = z'x$ for every z, which implies that $A'Ax = x$. But x is arbitrary, so $A'A = I$, that is, A is orthogonal. ∎

22.3.3 Lemma

$$\|x+y\|^2 - \|x-y\|^2 = 4x'y.$$

Proof:

$$\begin{aligned}\|x+y\|^2 - \|x-y\|^2 &= (x+y)'(x+y) - (x-y)'(x-y) \\ &= x'x + 2x'y + y'y - \big(x'x - 2x'y + y'y\big) \\ &= 4x'y.\end{aligned}$$

∎

22.3.1 Quadratic forms

An expression of the form

$$x'Ax,$$

where x is an $n \times 1$ column vector and A is an $n \times n$ matrix, is called a **quadratic form**[1] in x, and

$$x'Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j.$$

If A and B are $n \times n$ and x, y are n-vectors, then

$$x'(A+B)y = x'Ay + x'By \quad \text{and} \quad (x+y)'A(x+y) = x'Ax + 2x'Ay + y'Ay.$$

The quadratic form, or the matrix A, is called **positive definite** if

$$x'Ax > 0 \text{ whenever } x \neq 0,$$

and **positive semidefinite** if

$$x'Ax \geqslant 0 \text{ whenever } x \neq 0.$$

Letting x be the i^{th} unit coordinate vector, we have $x'Ax = a_{ii}$. As a consequence,

- if A is positive definite, then the diagonal elements satisfy $a_{ii} > 0$.
- If A is only positive semidefinite, then $a_{ii} \geqslant 0$.
- The level surfaces of a positive definite quadratic form are ellipsoids in $\boldsymbol{R}^{\mathrm{n}}$.

If A is of the form $A = B'B$, then A is necessarily positive semidefinite, since $x'Ax = x'B'Bx = (Bx)'(Bx) = \|Bx\|^2$. This also implies that if A is positive semidefinite and nonsingular, then it is in fact positive definite. There is also a converse. If A is positive semidefinite, then there is a square matrix B such that $A = B'B$ (sort of like a square root of A).

[1] For decades I was baffled by the term *form*. I once asked Tom Apostol at a faculty cocktail party what it meant. He professed not to know (it was a cocktail party, so that is excusable), but suggested that I should ask John Todd. He hypothesized that mathematicians don't know the difference between form and function, a clever reference to modern architectural philosophy. Professor Todd was quite imperious looking and had an English accent so I was too intimidated to ask him, but I subsequently learned (where, I can't recall) that *form* refers to a polynomial function in several variables where each term in the polynomial has the same degree. (The *degree* of the term is the sum of the exponents. For example, in the expression $xyz + x^2y + xz + z$, the first two terms have degree three, the third term has degree two and the last one has degree one. It is thus not a form.) This is most often encountered in the phrases *linear form* (each term has degree one) or *quadratic form* (each term has degree two).

22.3.2 Gradients of linear and quadratic forms

A function f given by

$$f(x) = a'Ax,$$

where a is $m \times 1$, A is $m \times n$ and x is $n \times 1$ is a linear function of x so the partial derivative $\partial f(x)/\partial x_j$ is just the coefficient on x_j, which is $a'A^j$. Thus, the gradient is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = A'a.$$

For a symmetric square $n \times n$ matrix, the quadratic form

$$Q(x) = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

satisfies $\partial Q(x)/\partial x_k = \sum_{i=1}^n a_{ik} x_i + \sum_{j=1}^n a_{kj} x_j = 2 \sum_{\ell=1}^n a_{k\ell} x_\ell$, where the last equality is due to the symmetry of A. Thus

$$\nabla Q(x) = \begin{bmatrix} \frac{\partial Q(x)}{\partial x_1} \\ \vdots \\ \frac{\partial Q(x)}{\partial x_n} \end{bmatrix} = 2Ax.$$

22.4 Eigenthingies and quadratic forms

If A is an $n \times n$ symmetric matrix, if there is a nonzero vector x and real number λ so that

$$Ax = \lambda x,$$

then x is called an **eigenvector** of A and λ is its corresponding **eigenvalue**.

- For a given eigenvalue λ, the set of corresponding eigenvectors together with zero form a linear subspace of $\boldsymbol{R}^{\mathrm{n}}$ called the **eigenspace** of λ.
- The dimension of this subspace is the **multiplicity** of λ.
- The eigenspaces of distinct eigenvalues are orthogonal.
- The sum of the multiplicities of the eigenvalues of A is n, or equivalently,
- there is an orthogonal basis for $\boldsymbol{R}^{\mathrm{n}}$ consisting of eigenvectors of A.

The Principal Axis Theorem may also be known as the Spectral Theorem, or the Diagonalization Theorem. It says that for a symmetric matrix, we can rotate the coordinate system so that the transformation $x \mapsto Ax$ simply multiplies each new coordinate j of x by some λ_j.

22.4.1 Principal Axis Theorem *Let A be an $n \times n$ symmetric matrix. Let $x_1, \dots, x_n$ be an orthonormal basis for $\boldsymbol{R}^{\mathrm{n}}$ made up of eigenvectors of A, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Set*

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix},$$

and let C be the matrix whose columns are $x_1, \dots, x_n$.

Then

$$A = C\Lambda C^{-1},$$
$$\Lambda = C^{-1}AC,$$

and C *is orthogonal, that is,*

$$C^{-1} = C'.$$

We can use this to **diagonalize** the quadratic form A. Let C and Λ be as in the Principal Axis Theorem. Given x, let

$$y = C'x = C^{-1}x, \text{ so } x = Cy.$$

Then

$$x'Ax = (Cy)'A(Cy) = y'C'ACy = y'\Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2.$$

22.5 Orthogonal projection onto a subspace

Let M be a linear subspace of $\boldsymbol{R}^{\mathrm{n}}$. The set of vectors in $\boldsymbol{R}^{\mathrm{n}}$ that are orthogonal to every $x \in M$ is a linear subspace of $\boldsymbol{R}^{\mathrm{n}}$, and it is called the **orthogonal complement** of M, denoted $M_\perp$

22.5.1 Orthogonal Complement Theorem *For each* $x \in \boldsymbol{R}^{\mathrm{n}}$ *we can write* x *in a unique way as* $x = x_M + x_\perp$*, where* $x_M \in M$ *and* $x_\perp \in M_\perp$*. The vector* x_M *is called the* ***orthogonal projection*** *of* x *onto* M*.*

In fact, let $y_1, \dots, y_m$ be an orthonormal basis for M. Put $z_i = (x \cdot y_i)y_i$ for $i = 1, \dots, m$. Then $x_M = \sum_{i=1}^m z_i$.

A key property of the orthogonal projection of x on M is that x_M is the point in M closest to x.

Formally, we have:

22.5.2 Proposition *Let* M *be a linear subspace of* $\boldsymbol{R}^{\mathrm{n}}$*. Let* $y \in \boldsymbol{R}^{\mathrm{n}}$*. Then*

$$\|y - y_M\| \leqslant \|y - x\| \quad \text{for all } x \in M.$$

Proof: This is really just the Pythagorean Theorem. Let $x \in M$. Then

$$\begin{aligned}
\|y - x\|^2 &= \|(y - y_M) + (y_M - x)\|^2 \\
&= \Big((y - y_M) + (y_M - x)\Big) \cdot \Big((y - y_M) + (y_M - x)\Big) \\
&= (y - y_M) \cdot (y - y_M) + 2(y - y_M) \cdot (y_M - x) + (y_M - x) \cdot (y_M - x)
\end{aligned}$$

but $y - y_M = y_\perp \perp M$, and $y_M - x \in M$, so $(y - y_M) \cdot (y_M - x) = 0$, so

$$\begin{aligned}
&= \|y - y_M\|^2 + \|y_M - x\|^2 \\
&\geqslant \|y - y_M\|^2.
\end{aligned}$$

Note that x, y_M, and y form a right triangle, so this is the Pythagorean Theorem.

That is, $x = y_M$ minimizes $\|y - x\|$ over M. ∎

22.5.3 Proposition (Linearity of Projection) *The orthogonal projection satisfies*

$$(x + z)_M = x_M + z_M \quad \text{and} \quad (\alpha x)_M = \alpha x_M.$$

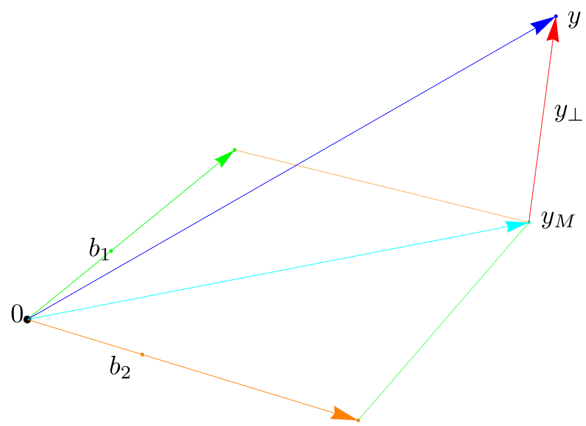

Figure 22.2. Orthogonal projection onto $M = \operatorname{span}\{b_2, b_2\}$.

Proof: Let $y_1, \dots, y_k$ be an orthonormal basis for M. Use $x_M = \sum_{j=1}^k (x \cdot y_j) y_j$ and $z_m = \sum_{j=1}^k (x \cdot z_j) z_j$. Then

$$(x + z)_M = \sum_{j=1}^k (x + z \cdot y_j) y_j.$$

Use linearity of the dot product. ∎

Since orthogonal projection onto M is a linear transformation, there is a matrix P such that for every x, $x_M = Px$. What does it look like? Let $M \subset \boldsymbol{R}^{\mathrm{n}}$ be a k dimensional subspace.

Let $x_1, \dots, x_k$ be a basis for M, and let X be the $n \times k$ matrix whose columns are $x_1, \dots, x_k$, and set

$$P = X(X'X)^{-1}X'.$$

Then for any $y \in \boldsymbol{R}^{\mathrm{n}}$,

$$y_M = Py = X(X'X)^{-1}X'y.$$

There are a number of points worth mentioning.

- Note that P is symmetric and **idempotent**, that is, $PP = P$.

- This implies that $X'X$ has an inverse. To see why, suppose $X'Xz = 0$. Then $z'X'Xz = 0$, but $z'X'Xz = \|Xz\|^2$ so $Xz = 0$. But Xz is a linear combination of the columns of X, which are independent by hypothesis, so $z = 0$. We have just shown that $X'Xz = 0 \implies z = 0$, so $X'X$ is invertible.

- Suppose y belongs to M. Then y is a linear combination of $x_1, \dots, x_k$ so there is a k-vector b with $y = Xb$. Then $Py = X(X'X)^{-1}X'Xb = Xb = y$, so P acts as the identity on M and every nonzero vector in M is an eigenvector of P corresponding to an eigenvalue of 1.

- Now suppose that y is orthogonal to M. Then $y'X = 0$, which implies $X'y = 0$, so $Py = X(X'X)^{-1}X'y = 0$. Thus every nonzero vector in $M_\perp$ is an eigenvector of P corresponding to an eigenvalue 0.

- We now show that $y - Py$ is orthogonal to every basis vector x_j. That is, we want to show that $(y - Py)'X = 0$. Now

$$(y - Py)'X = y'X - y'P'X = y'X - y'X'(X'X)^{-1}X'X = y'X - y'X = 0.$$

By the uniqueness of the orthogonal decomposition this shows that P is the orthogonal projection onto M.

- Thus Py is always a linear combination $x_1, \dots, x_k$, that is, $Py = Xb$ for some k-vector b, namely

$$b = (X'X)^{-1}X'y.$$

In fact, we have just shown that any symmetric idempotent matrix is a projection on its range.

22.6 Review of random vectors

Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

be a random vector. Define

$$\boldsymbol{\mu} = \boldsymbol{E}\,\boldsymbol{X} = \begin{bmatrix} \boldsymbol{E}\,X_1 \\ \vdots \\ \boldsymbol{E}\,X_n \end{bmatrix}.$$

Define the **covariance matrix** of $\boldsymbol{X}$ by

$$\boldsymbol{Var}\,\boldsymbol{X} = \big[\boldsymbol{Cov}\,X_i X_j\big] = \big[\boldsymbol{E}(X_i - \boldsymbol{\mu}_i)(X_j - \boldsymbol{\mu}_j)\big] = \big[\sigma_{ij}\big] = \boldsymbol{E}\left((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\right).$$

Let

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

be an $m \times n$ matrix of constants, and let

$$\boldsymbol{Y} = A\boldsymbol{X}$$

Then, since expectation is a linear operator,

$$\boldsymbol{E}\,\boldsymbol{Y} = A\boldsymbol{\mu}.$$

Moreover

$$\boldsymbol{Var}\,\boldsymbol{Y} = A(\,\boldsymbol{Var}\,\boldsymbol{X})A'$$

since $\boldsymbol{Var}\,\boldsymbol{Y} = \boldsymbol{E}\left((A\boldsymbol{X} - A\boldsymbol{\mu})(A\boldsymbol{X} - A\boldsymbol{\mu})'\right) = \boldsymbol{E}\left(A(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'A'\right) = A(\,\boldsymbol{Var}\,\boldsymbol{X})A'$.

The covariance matrix Σ of a random vector $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ is always positive semidefinite, since for any vector w of weights, $w'\Sigma w$ is the variance of the random variable $w'Y$, and variances are always nonnegative.

22.7 The Normal density

Recall that the Normal $N(\mu, \sigma^2)$ has a density of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}(x-\mu)\frac{1}{\sigma^2}(x-\mu)} \tag{2}$$

although I don't normally write it in that form.

22.8 The Multivariate Normal distribution

I am going to use a not quite standard definition of the multivariate normal that makes life simpler. I learned this approach from Dave Grether who told me that he learned it from the late statistician Edwin James George Pitman, who happens to be the father of Jim Pitman, the author of the probability textbook [9] for the course. I have also found this definition is the one used by Jacod and Protter [7, Definition 16.1, p. 126] and by Rao [10, p. 518].

22.8.1 Definition *Extend the notion of a Normal random variable to include constants as* $N(\mu, 0)$ *zero-variance random variables.*

A random vector $\boldsymbol{X} = (X_1, \dots, X_n) \in \boldsymbol{R}^{\mathrm{n}}$ *has a* ***multivariate Normal distribution*** *or a* ***jointly Normal distribution*** *if for every constant vector* $\boldsymbol{T} \in \boldsymbol{R}^{\mathrm{n}}$ *the linear combination* $\boldsymbol{T}'\boldsymbol{X} = \sum_{i=1}^n T_i X_i$ *has a Normal*$(\mu_{\boldsymbol{T}}, \sigma^2_{\boldsymbol{T}})$ *distribution.*

Recall Proposition 10.8.1:

22.8.2 Proposition *If* X *and* Y *are independent normals with* $X \sim N(\mu, \sigma^2)$ *and* $Y \sim N(\lambda, \tau^2)$*, then*

$$X + Y \sim N(\mu + \lambda, \sigma^2 + \tau^2).$$

It has the following Corollary:

22.8.3 Corollary *If* $X_1, \dots, X_n$ *are independent Normal random variables, then the random vector*

$$\boldsymbol{X} = (X_1, \dots, X_n)$$

has a multivariate Normal distribution.

Proof: We need to show that for any constant vector $\boldsymbol{T}$, the linear combination $\boldsymbol{T}'\boldsymbol{X} = \sum_{i=1}^n T_i X_i$ has a Normal distribution. But since the X_is are independent Normals, the $T_i X_i$s are also independent Normals, so by the Proposition, their sum is a Normal random variable. ∎

So now we know that multivariate Normal random vectors do exist.

22.8.4 Proposition *If* $\boldsymbol{X}$ *is an* n*-dimensional multivariate Normal random vector, and* A *is an* $m \times m$ *constant matrix, then*

$$\boldsymbol{Y} = A\boldsymbol{X}$$

is an m*-dimensional multivariate Normal random vector.*

Proof: For a constant $1 \times m$-vector $\boldsymbol{T}$, the linear combination $\boldsymbol{T}Y$ is just the linear combination $(\boldsymbol{T}A)\boldsymbol{X}$, which by hypothesis is Normal. ∎

22.8.5 Proposition *If* $\boldsymbol{X} = (X_1, \dots, X_n)$ *has a multivariate Normal distribution, then*

- *Each component* X_i *has a Normal distribution.*
- *Every subvector of* $\boldsymbol{X}$ *has a multivariate Normal distribution.*

22.8.6 Definition *Let* $\boldsymbol{\mu} = \boldsymbol{E}\,\boldsymbol{X}$ *and let*

$$\Sigma = \boldsymbol{E}\left((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\right),$$

that is,

$$\sigma_{i,j} := \Sigma_{i,j} = \boldsymbol{Cov}(X_i, X_j) = \boldsymbol{E}(X_i - \boldsymbol{E}\,X_i)(X_j - \boldsymbol{E}\,X_j)$$

22.8.7 Theorem *If $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, then*

$$\boldsymbol{Y} = C\boldsymbol{X} \sim N(C\boldsymbol{\mu}, C\Sigma C').$$

22.8.8 Corollary *Uncorrelated jointly normal random variables are in fact independent!!*

Proof: If the random variables are uncorrelated, then Σ is diagonal. In that case the quadratic form $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ reduces to a sum of squares, so the density factors into the product of the marginal densities, which implies independence. ∎

- Let C be diagonal. Then $C\boldsymbol{X}$ is a linear combination $c_1X_1 + \cdots + c_nX_n$ of the components and has a (univariate) normal $N(C\boldsymbol{\mu}, C\Sigma C')$ distribution.

- Σ is positive semi-definite.

To see this, let $\boldsymbol{a} \in \boldsymbol{R}^n$. Then $\boldsymbol{Var}(\boldsymbol{a}'\boldsymbol{X}) = \boldsymbol{a}'\Sigma\boldsymbol{a}$, but variances are always nonnegative.

The following theorem may be found in Jacod and Protter [7, Theorem 16.1, p. 126–127]. It is used by Anderson [1, § 2.4] as the definition of a multivariate normal.

22.8.9 Proposition *A random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ with nonsingular covariance matrix Σ has a multivariate normal distribution if and only if its density is*

$$f(\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det \Sigma}}\, e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{3}$$

Note that this agrees with (2) when $n = 1$.

- The multivariate Normal density is constant on ellipsoids of the form

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \text{constant}.$$

Singular covariance matrices

When a jointly Normal distribution has a singular covariance matrix, then the density does not exist. But if the matrix has rank k, there is a k-dimensional flat[2] on which it is possible to define a density of the form given in (3), but we will not make use of that here.

22.9 ⋆ Not jointly normal vectors

If $X_1, \ldots, X_n$ are independent Normals, then the vector $(X_1, \ldots, X_n)$ is a jointly Normal vector. But if the components are not independent, then the vector may not be jointly Normal. That is, there may be a linear combination of them that does not have a Normal distribution. Here are a couple of examples, the details of which I leave as an exercise.

22.9.1 Example (Example 2, [7, p. 134]) Let Z be a standard Normal and let $a > 0$. Define

$$X = \begin{cases} Z & \text{if } |Z| \leqslant a \\ -Z & \text{if } |Z| > a. \end{cases}$$

Then you can show that X has a standard Normal distribution, $X + Z$ is not constant, but $P(X + Z > 2a) = 0$. Thus (X, Z) is not jointly Normal. □

[2] A **flat** is a translate of a linear subspace. That is, if M is a linear subspace, then $x + M = \{x + y : y \in M\}$ is a flat.

22.9.2 Example (Exercise 4.47, [6, p. 200]) Let X and Y be independent standard normal random variables. Define the random variable Z by

$$Z = \begin{cases} X & \text{if } XY > 0 \\ -X & \text{if } XY < 0. \end{cases}$$

Then Z is a standard normal random variable, but the vector (Y, Z) is *not* jointly normal.

I leave the verification as an exercise. (Hint for the second claim: Y and Z always have the same sign.) □

See Casella and Berger [6, Exercises 4.48–4.50, pp. 200–201] for more peculiar examples of pairs of normals that are not jointly normal. Behboodian [3] describes a pair of uncorrelated normal random variables that are not independent—they aren't jointly normal either.

22.10 ⋆ Multivariate Central Limit Theorem

The following vector version of the Central Limit Theorem may be found in Jacod and Protter [7, Theorem 21.3, p. 183].

22.10.1 Multivariate Central Limit Theorem *Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_i \dots$ be independent and identically distributed random k-vectors with common mean $\boldsymbol{\mu}$ and covariance matrix Σ. Let $\boldsymbol{S}_n = \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n$. Then*

$$\frac{\boldsymbol{S}_n - \boldsymbol{\mu}}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \Sigma).$$

Note that this result asserts that the limiting distribution is jointly Normal, which is stronger than saying that the distribution of each component is Normal.

22.11 Multivariate Normal and Chi-square

22.11.1 Proposition *Let $\boldsymbol{X} \sim N(0, I_n)$. Then $\boldsymbol{X}'A\boldsymbol{X} \sim \chi^2(k)$ if and only if A is symmetric, idempotent, and has rank k.*

Proof: I'll prove only one half. Assume A is symmetric, idempotent, and has rank k. (Then it is orthogonal projection onto a k-dimensional subspace.) Its eigenvalues are 0 and 1, so it is positive semidefinite. So by the Principal Axis Theorem, there is an orthogonal matrix C such that

$$C'AC = \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

where the λ_is are the eigenvalues, k of them are 1 and $n - k$ are zero. Setting $\boldsymbol{Y} = C'\boldsymbol{X}$, we see that $\boldsymbol{Y} \sim N(0, \Lambda)$. This means that the components of $\boldsymbol{Y}$ are independent. Moreover the Principal Axis Theorem also implies

$$\boldsymbol{X}'A\boldsymbol{X} = \boldsymbol{Y}'\Lambda\boldsymbol{Y} = \sum_{i:\lambda_i=1} Y_i^2,$$

which has a $\chi^2(k)$ distribution. ∎

22.11.2 Corollary *Let $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \sigma^2 I_n)$. Then*

$$\left(\frac{\boldsymbol{X} - \boldsymbol{\mu}}{\sigma}\right)' A \left(\frac{\boldsymbol{X} - \boldsymbol{\mu}}{\sigma}\right) \sim \chi^2(k)$$

if and only if A is symmetric, idempotent, and has rank k.

Proof: Note that $\frac{\boldsymbol{X}-\boldsymbol{\mu}}{\sigma} \sim N(0, I)$. ∎

The following theorem is useful in deriving the distribution of certain test statistics. You can find a proof in C. Radakrishna Rao [10, Item (v), p. 187] or Henri Theil [11, pp. 83–84].

22.11.3 Theorem *Let $\boldsymbol{X} \sim N(0, \sigma^2 I)$ and let A_1 and A_2 be symmetric idempotent matrices that satisfy*

$$A_1 A_2 = A_2 A_1 = 0.$$

Then $\boldsymbol{X}' A_1 \boldsymbol{X}$ and $\boldsymbol{X}' A_2 \boldsymbol{X}$ are independent.

22.12 Independence of Sample Mean and Variance Statistics

Let $X_1, \dots, X_n$ be independent and identically distributed Normal $N(\mu, \sigma^2)$ random variables. Let $D_i = X_i - \bar{X}$.

22.12.1 Theorem *The sample average $\bar{X}$ and the random vector $(D_1, \dots, D_n)$ are stochastically independent. [N.B. This does not say that each D_i and D_j are independent of each other, rather, they are jointly independent of $\bar{X}$.]*

Proof: The random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ has a multivariate Normal distribution (Corollary 22.8.3). Therefore the random vector

$$\begin{bmatrix} D_1 \\ \vdots \\ \vdots \\ D_n \\ \bar{X} \end{bmatrix} = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & & \ddots & & \vdots \\ & & & & \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & & 1-\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & & \frac{1}{n} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ \vdots \\ X_n \end{bmatrix}$$

is also a multivariate Normal random vector (Theorem 22.8.7).

But by the Covariance Menagerie 9.8.1.15, $\boldsymbol{Cov}(D_i, \bar{X}) = 0$. But for multivariate Normal vectors, this means that S and $(D_1, \dots, D_n)$ are stochastically independent. ∎

22.12.2 Corollary *If $X_1, \dots, X_n$ are independent and identically distributed Normal $N(\mu, \sigma^2)$ random variables. Define*

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

1. $\bar{X} \sim N(\mu, \sigma^2/n)$.

2. *$\bar{X}$ and S^2 are independent.*

3. $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$

Proof: (1). This is old news.

(2). By Theorem 22.12.1, $\bar{X}$ is independent of $(D_1, \dots, D_n)$. But $S^2 = \sum_i D_i^2/(n-1)$ is a function of $(D_1, \dots, D_n)$, so it too is independent of $\bar{X}$.

(3). Define the standardized version of X_i,

$$Y_i = \frac{X_i - \mu}{\sigma}, \quad (i = 1, \dots, n), \quad \bar{Y} = \sum_{i=1}^n Y_i/n = (\bar{X} - \mu)/\sigma.$$

Note that for each i,

$$Y_i - \bar{Y} = (X_i - \bar{X})/\sigma,$$

each Y_i is a standard Normal random variable, and the Y_is are independent. So $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ multivariate Normal with covariance matrix I. Let

$$\boldsymbol{v} = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right).$$

Note that $\boldsymbol{v}'\boldsymbol{v} = 1$. Now create an orthogonal matrix B that has $\boldsymbol{v}$ as its last row. We can always do this. (Think of the Gram–Schmidt procedure.)

Define the transformed variables

$$\boldsymbol{Z} = B\boldsymbol{Y}.$$

Since $\boldsymbol{Y}$ is multivariate Normal, therefore so is $\boldsymbol{Z}$. By Proposition 22.8.4, the covariance matrices satisfy

$$\boldsymbol{Var}\,\boldsymbol{Z} = B(\boldsymbol{Var}\,\boldsymbol{Y})B' = BB' = BB^{-1} = I,$$

so $\boldsymbol{Z}$ is a vector of independent standard Normal random variables.

By Proposition 22.3.2, multiplication by B preserves inner products, so

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} Y_i^2. \tag{4}$$

But Z_n is the dot product of the last row of B with $\boldsymbol{Y}$, or

$$Z_n = \boldsymbol{v} \cdot \boldsymbol{Y} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = \sqrt{n}\bar{Y}. \tag{5}$$

So combining (4) and (5), we have

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n-1} Z_i^2 + n\bar{Y}^2. \tag{6}$$

On the other hand, we can write

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + n\bar{Y}^2, \tag{7}$$

Combining (6) and (7) implies

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n-1} Z_i^2. \tag{8}$$

Now rewrite this in terms of the X_is:

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2. \tag{9}$$

Combining (8) and (9) shows that

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n-1} Z_i^2$$

where Z_i are independent standard Normals. In other words, $\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2$ has a $\chi^2(n-1)$ distribution. ∎

Bibliography

[1] T. W. Anderson. 1958. *An introduction to multivariate statistical analysis*. A Wiley Publication in Mathematical Statistics. New York: Wiley.

[2] S. Axler. 1997. *Linear algebra done right*, 2d. ed. Undergraduate Texts in Mathematics. New York: Springer.

[3] J. Behboodian. 1990. Examples of uncorrelated dependent random variables using a bivariate mixture. *American Statistician* 44(3):218. http://www.jstor.org/stable/2685339

[4] K. C. Border. More than you wanted to know about quadratic forms. http://www.hss.caltech.edu/~kcb/Notes/QuadraticForms.pdf

[5] ——— . Quick review of matrix and linear algebra. http://www.hss.caltech.edu/~kcb/Notes/LinearAlgebra.pdf

[6] G. Casella and R. L. Berger. 2002. *Statistical inference*, 2d. ed. Pacific Grove, California: Wadsworth.

[7] J. Jacod and P. Protter. 2004. *Probability essentials*, 2d. ed. Berlin and Heidelberg: Springer.

[8] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[9] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[10] C. R. Rao. 1973. *Linear statistical inference and its applications*, 2d. ed. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

[11] H. Theil. 1971. *Principles of econometrics*. New York: Wiley.