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

Relevant textbook passages:
Larsen–Marx [6]: Section 10.5.
Pitman [7]: Section 6.5.

22.1 News flash

It was announced on February 12, 2015, that the journal Basic and Applied Social Psychology (BASP) has banned the use of the null hypothesis significance testing procedure (NHSTP) in submissions to the journal, on the grounds that the procedures have been abused and no longer serve their intended purpose. See the editorial at the journal’s web site and the commentary here.

22.2 A reminder of some linear and matrix algebra

If you just need a quick refresher, I recommend my on-line notes [3, 4]. In what follows, vectors in $\mathbb{R}^n$ are usually considered to be $n \times 1$ column matrices, and $'$ denotes transposition. To save space I may write a vector $a = (x_1, \ldots, x_n)$, but you should still think of it as a column vector.

22.2.1 The geometry of the Euclidean inner product

Vectors $x$ and $y$ in $\mathbb{R}^n$ 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 $\theta$ 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

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

Referring to Figure 22.1 we see that

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

22–1
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,

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

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 $x$,

$$\|Ax\| = \|x\|.$$  \hfill (1)

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

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

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,

$$(Ax) \cdot (Ay) = \frac{\|Ax + Ay\| - \|Ax - Ay\|}{4}$$

$$= \frac{\|A(x + y)\| - \|A(x - y)\|}{4}$$

$$= \frac{\|x + y\| - \|x - y\|}{4}$$

$$= x'y.$$ 

(3) $\implies$ (1) Assume $A$ preserves inner products. Pick an arbitrary $x$ and $z$, and let $y = Az$. Then $y'Ax = z'A'Ax = z'x$ 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. \hfill \blacksquare

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

v. 2015.02.28::16.30

KC Border
22.3.3 Lemma

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

Proof:

\[
\|x + y\|^2 - \|x - y\|^2 = (x + y)'(x + y) - (x - y)'(x - y) \\
= x'x + 2x'y + y'y - (x'x - 2x'y + y'y) \\
= 4x'y.
\]

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

\[ x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_ix_j. \]

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

\[ x'Ax > 0 \quad \text{whenever} \quad x \neq 0, \]

and positive semidefinite if

\[ x'Ax \geq 0 \quad \text{whenever} \quad x \neq 0. \]

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 definite semidefinite, then \( a_{ii} \geq 0 \).

---

\(^1\)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. I was too intimidated by Todd to ask, 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 \( x_1x_2 + x_2^2y + x_2z + z \), 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).
• The level surfaces of a positive definite quadratic form are ellipsoids in $\mathbb{R}^n$.

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 non-singular, 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 $A$).

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

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

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 $\lambda$ so that

$$Ax = \lambda x,$$

then $x$ is called an eigenvector of $A$ and $\lambda$ is its corresponding eigenvalue.

• For a given eigenvalue $\lambda$, the set of corresponding eigenvectors together with zero form a linear subspace of $\mathbb{R}^n$ called the eigenspace of $\lambda$.

• The dimension of this subspace is the multiplicity of $\lambda$.

• 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 $\mathbb{R}^n$ consisting of eigenvectors of $A$.

22.4.1 Definition A square matrix $A$ is orthogonal if $A'A = I$, or equivalently $A' = A^{-1}$.

The condition $A'A = I$ means that the columns of $A$ are orthonormal (orthogonal and norm 1), and so are the rows.

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

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

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

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 $\Lambda$ be as in the Principal Axis Theorem. Given $x$, let

$$y = C'x = C^{-1}x,$$ 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 $\mathbb{R}^n$. The set of vectors in $\mathbb{R}^n$ that are orthogonal to every $x \in M$ is a linear subspace of $\mathbb{R}^n$, and it is called the orthogonal complement of $M$, denoted $M_{\perp}$.

22.5.1 Orthogonal Complement Theorem For each $x \in \mathbb{R}^n$ 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 $y_1, \ldots, y_m$ be an orthonormal basis for $M$. Put $z_i = (x \cdot y_i)y_i$ for $i = 1, \ldots, m$. 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 $x$. 

KC Border v. 2015.02.28::16.30
Formally, we have:

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

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

*Proof:* This is really just the Pythagorean Theorem is disguise. Let $x \in \mathbb{R}^n$. Then

\[
\|y - x\|^2 = \|(y - y_M) + (y_M - x)\|^2 \\
= ((y - y_M) + (y_M - x)) \cdot ((y - y_M) + (y_M - x)) \\
= (y - y_M) \cdot (y - y_M) + 2(y - y_M) \cdot (y_M - x) + (y_M - x) \cdot (y_M - x)
\]

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

\[
= \|y - y_M\|^2 + \|y_M - x\|^2 \\
\geq \|y - y_M\|^2.
\]

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

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

*Proof:* Let $y_1, \ldots, y_k$ 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 $M \subset \mathbb{R}^n$ be a $k$ dimensional subspace.

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

\[
P = X(X'X)^{-1}X'.
\]

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

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

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 $x_1, \ldots, x_k$ 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

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

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

Thus $Py$ is always a linear combination $x_1, \ldots, x_k$, 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

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

be a random vector. Define

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

Define the **variance-covariance matrix** of $X$ by

$$\text{Var} X = \text{Cov} X_i X_j = \begin{bmatrix} E(X_i - \mu_i)(X_j - \mu_j) \end{bmatrix} = \begin{bmatrix} \sigma_{ij} \end{bmatrix} = E((X - \mu)(X - \mu)') .$$
Let
\[ A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \]
be an \( m \times n \) matrix of constants, and let
\[ Y = AX \]
Then, since expectation is a linear operator,
\[ EY = A\mu. \]
Moreover
\[ \text{Var} Y = A(\text{Var} X)A' \]
since
\[ \text{Var} Y = E\left((AX - A\mu)(AX - A\mu)'\right) = E\left(A(X - \mu)(X - \mu)'A'\right) = A(\text{Var} X)A'. \]
The variance-covariance matrix \( \Sigma \) of a random vector \( Y = (Y_1, \ldots, Y_n) \) 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
\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2} \]
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 \([7]\) for the course. I have also found this definition is the one used by Jacod and Protter \([5]\), Definition 16.1, p. 126]. You should consult Anderson \([1, \S\ 2.4]\) to verify that this agrees with the usual definitions.

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 \( X = (X_1, \ldots, X_n) \in \mathbb{R}^n \) has a \textbf{multivariate Normal distribution} or a \textbf{jointly Normal distribution} if for every constant vector \( T \in \mathbb{R}^n \) the linear combination \( T'X = \sum_{i=1}^n T_iX_i \) has a \( \text{Normal}(\mu_T, \sigma_T^2) \) distribution.
Recall Proposition 9.15.1:

**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 $X_1, \ldots, X_n$ are independent Normal random variables, then the random vector

$$X = (X_1, \ldots, X_n)$$

has a multivariate Normal distribution.

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

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

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

$$Y = AX$$

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

*Proof*: For a constant $1 \times m$-vector $T$, the linear combination $TY$ is just the linear combination $(TA)X$, which by hypothesis is Normal. $\blacksquare$

**22.8.5 Proposition** If $X = (X_1, \ldots, X_n)$ has a multivariate Normal distribution, then

- Each component $X_i$ has a Normal distribution.
- Every subvector of $X$ has a multivariate Normal distribution.

**22.8.6 Definition** Let $\mu = E X$ and let

$$\Sigma = E ((X - \mu)(X - \mu)'),$$

that is,

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

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

$$Y = CX \sim N(C\mu, C\Sigma C').$$
22.8.8 Corollary Uncorrelated jointly normal random variables are in fact independent!!

Proof: If the random variables are uncorrelated, then $\Sigma$ is diagonal. In that case the quadratic form $(x - \mu)'\Sigma^{-1}(x - \mu)$ 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 $CX$ is a linear combination $c_1X_1 + \cdots + c_nX_n$ of the components and has a (univariate) normal $N(C\mu, C\Sigma C')$ distribution.

- $\Sigma$ is positive semi-definite.

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

The following theorem may be found in Anderson [1, Theorems 2.3.1, p. 17].

22.8.9 Proposition If $X = (X_1, \ldots, X_n)$ has a multivariate normal distribution, and if the variance-covariance matrix $\Sigma$ is nonsingular, then its density is

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

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

- The multivariate Normal density is constant on ellipsoids of the form $(x - \mu)'\Sigma^{-1}(x - \mu) = \text{constant}$.

22.9 Multivariate Normal and Chi-square

22.9.1 Proposition Let $X \sim N(0, I_n)$. Then $X'AX \sim \chi^2(k)$ 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

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

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

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

which has a $\chi^2(k)$ distribution.
22.9.2 Corollary Let \( X \sim N(\mu, \sigma^2 I_n) \). Then
\[
\left( \frac{X - \mu}{\sigma} \right)' A \left( \frac{X - \mu}{\sigma} \right) \sim \chi^2(k)
\]
if and only if \( A \) is symmetric, idempotent, and has rank \( k \).

Proof: Note that \( X - \mu \sim N(0, I) \).

22.9.3 Theorem Let \( X \sim N(0, \sigma^2 I_n) \) and let \( A_1 \) and \( A_2 \) be symmetric idempotent matrices that satisfy
\[
A_1 A_2 = A_2 A_1 = 0.
\]
Then \( X'A_1X \) and \( X'A_2X \) are independent.

22.10 A covariance menagerie

Here is a result that should have come in Lecture 9. Recall that for independent random variables \( X \) and \( Y \), \( \text{Var}(X + Y) = \text{Var}X + \text{Var}Y \), and \( \text{Cov}(XY) = 0 \). For any random variable \( X \) with finite variance, \( \text{Var}X = E(X^2) - (E X)^2 \), so \( E(X^2) = \text{Var}X + (E X)^2 \). If \( E X = 0 \), then \( \text{Cov}(XY) = E(XY) \).

Let \( X_1, \ldots, X_n \) be independent and identically distributed random variables with common mean \( \mu \) and variance \( \sigma^2 \). Define the total and averages
\[
T = \sum_{i=1}^n X_i, \quad \bar{X} = T/n,
\]
and let
\[
D_i = X_i - \bar{X}, \quad (i = 1, \ldots, n).
\]
be the deviation of \( X_i \) from \( \bar{X} \).

22.10.1 Theorem (A Covariance Menagerie) We have:

1. \( E(X_iX_j) = (E X_i)(E X_j) = \mu^2 \), for \( i \neq j \) (by independence).
2. \( E(X_i^2) = \sigma^2 + \mu^2 \).
3. \( E(X_iT) = \sum_{j=1}^n E(X_iX_j) + E(X_i^2) = E(X_i^2) = \sigma^2 + n\mu^2 \).
4. \( E(X_i\bar{X}) = E(X_iT/n) = (\sigma^2/n) + \mu^2 \).
5. \( E(T) = n\mu \).
6. \( \text{Var}(T) = n\sigma^2 \).
7. \( E(T^2) = n\sigma^2 + n^2\mu^2 \).
8. \( E(\bar{X}) = \mu \).
9. \( \text{Var}(\bar{X}) = \sigma^2/n. \)
10. \( E(\bar{X}^2) = (\sigma^2/n) + \mu^2. \)
11. \( E(D_i) = 0, \ i = 1, \ldots, n. \)
12. \( \text{Var}(D_i) = E(D_i^2) = (n-1)\sigma^2/n \)
\[
\text{Var}(D_i) = E(X_i - \bar{X})^2 = E(X_i^2) - 2E(X_i\bar{X}) + E(\bar{X}^2) = \\
(\sigma^2 + \mu^2) - 2((\sigma^2/n) + \mu^2) + ((\sigma^2/n) + \mu^2) = \left(1 - \frac{1}{n}\right)\sigma^2.
\]
13. \( \text{Cov}(D_i, D_j) = E(D_iD_j) = -\sigma^2/n, \text{ for } i \neq j. \)
\[
E(D_iD_j) = E((X_i - \bar{X})(X_j - \bar{X})) = E(X_iX_j) - E(X_i\bar{X}) - E(X_j\bar{X}) + E(\bar{X}^2) = \\
\mu^2 - [(\sigma^2/n) + \mu^2] - [(\sigma^2/n) + \mu^2] + [(\sigma^2/n) + \mu^2] = -\sigma^2/n.
\]
14. \( \text{Cov}(D_i, T) = E(D_iT) = 0. \)
\[
E(D_iT) = E\left((X_i - (T/n))T\right) = E(X_iT) - E(T^2/n) = \\
(\sigma^2 + n\mu^2) - (n\sigma^2 + n^2\mu^2)/n = 0.
\]
15. \( \text{Cov}(D_i, \bar{X}) = E(D_i\bar{X}) = E(D_iT)/n = 0. \)

The proof of each is a straightforward plug-and-chug calculation. The only reason for writing this as a theorem is to be able to refer to it easily.

**22.11 Independence of Sample Mean and Variance Statistics**

Let \( X_1, \ldots, X_n \) be independent and identically distributed Normal \( N(\mu, \sigma^2) \) random variables.

**22.11.1 Theorem** Then using the terminology of the previous section, \( \bar{X} \) and the random vector \( (D_1, \ldots, D_n) \) 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 $\mathbf{X} = (X_1, \ldots, X_n)$ has a multivariate Normal distribution (Corollary 22.8.3). Therefore the random vector

$$(D_1, D_2, \ldots, D_n, \bar{X}) = \begin{bmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & -\frac{1}{n} & \cdots \\ & \ddots & \ddots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \mathbf{X}$$

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

But by the Covariance Menagerie 22.10.14, $\text{Cov}(D_i, \bar{X}) = 0$. But for multivariate Normal vectors, this means that $S$ and $(D_1, \ldots, D_n)$ are stochastically independent.

22.11.2 Corollary If $X_1, \ldots, X_n$ 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. $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2(n - 1)$

Proof: This proof is similar to Larsen–Marx [6, pp. 424–425].

(1). This is old news.

(2). By Theorem 22.11.1, $\bar{X}$ is independent of $(D_1, \ldots, D_n)$. But $S^2 = \sum_i D_i^2/(n - 1)$ is a function of $(D_1, \ldots, D_n)$, so it too is independent of $X$.

(3). Define the standardized version of $X_i$,

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

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 $Y_i$s are independent. So $\mathbf{Y} = (Y_1, \ldots, Y_n)$ multivariate Normal with variance-covariance matrix $I$. Let

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

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

\[ Z = BY. \]

Since \( Y \) is multivariate Normal, therefore so is \( Z \). By Proposition 22.8.4, the variance-covariance matrices satisfy

\[ \text{Var} \ Z = B(\text{Var} \ Y)B' = BB' = BB^{-1} = I, \]

so \( Z \) is a vector of independent standard Normal random variables.

By Proposition 22.3.2, multiplication by \( B \) preserves inner products, so

\[ \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} Y_i^2. \]  \hspace{1cm} (4)

But \( Z_n \) is the dot product of the last row of \( B \) with \( Y \), or

\[ Z_n = v \cdot Y = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = \sqrt{n} \bar{Y}. \]  \hspace{1cm} (5)

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

\[ \sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n-1} Z_i^2 + n \bar{Y}^2. \]  \hspace{1cm} (6)

On the other hand, we can write

\[ \sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + n \bar{Y}^2. \]  \hspace{1cm} (7)

Combining (6) and (7) implies

\[ \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n-1} Z_i^2. \]  \hspace{1cm} (8)

Now rewrite this in terms of the \( X_i \)s:

\[ \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2. \]  \hspace{1cm} (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.  \[ \blacksquare \]
Bibliography


   http://www.hss.caltech.edu/~kcb/Notes/QuadraticForms.pdf

   http://www.hss.caltech.edu/~kcb/Notes/LinearAlgebra.pdf


