Notes 1: Normal Distribution and Linear Algebra

Nathaniel E. Helwig

Department of Statistics
University of Illinois at Urbana-Champaign

Stat 420: Methods of Applied Statistics
Section N1U/N1G – Spring 2014

Outline of Notes

1) Univariate Normal:
- Distribution form
- Standard normal
- Probability calculations
- Affine transformations

2) Bivariate Normal:
- Distribution form
- Probability calculations
- Affine transformations
- Conditional distributions

3) Linear Algebra:
- Some basics
- Matrix definiteness
- Matrix decompositions
- Determinants and inverses

4) Multivariate Normal:
- Distribution form
- Probability calculations
- Affine transformations
- Conditional distributions

Normal Density Function (Univariate)

Given a variable $x \in \mathbb{R}$, the normal probability density function (pdf) is

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \end{aligned} \tag{1}$$

where

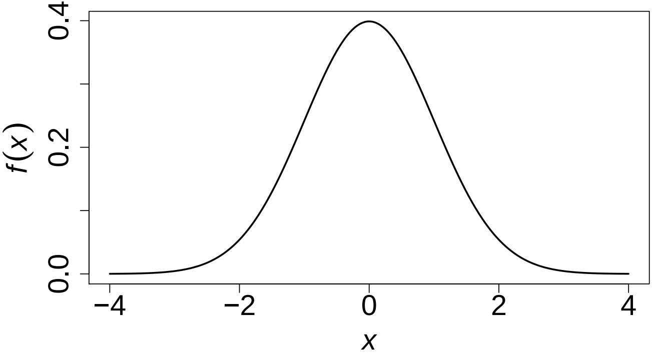
- $\mu \in \mathbb{R}$ is the mean
- $\sigma > 0$ is the standard deviation (σ^2 is the variance)
- $e \approx 2.71828$ is base of the natural logarithm

Write $X \sim \mathrm{N}(\mu, \sigma^2)$ to denote that X follows a normal distribution.

Standard Normal Distribution

If $X \sim \mathrm{N}(0, 1)$, then X follows a standard normal distribution:

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



Probabilities and Distribution Functions

Probabilities relate to the area under the pdf:

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f(x)\mathrm{d}x \\ &= F(b) - F(a) \end{aligned} \tag{3}$$

where

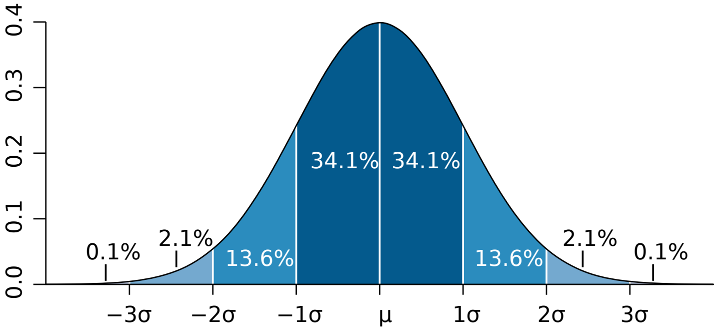
$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u \tag{4}$$

is the *cumulative distribution function* (cdf).

Note: $\quad F(x) = P(X \leq x) \quad \Longrightarrow \quad 0 \leq F(x) \leq 1$

Normal Probabilities

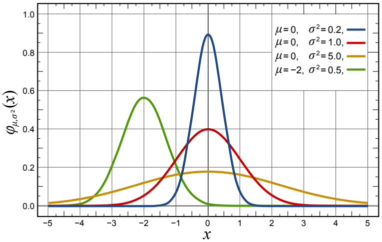
Helpful figure of normal probabilities:



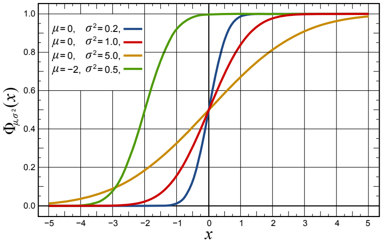
From http://en.wikipedia.org/wiki/File:Standard_deviation_diagram.svg

Normal Distribution Functions (Univarite)

Helpful figures of normal pdfs and cdfs:



http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg



http://en.wikipedia.org/wiki/File:Normal_Distribution_CDF.svg

Note that the cdf has an elongated "S" shape, referred to as an *ogive*.

Affine Transformations of Normal (Univariate)

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$ with $a \neq 0$.

If we define $Y = aX + b$, then $Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.

Suppose that $X \sim \mathrm{N}(1, 2)$. Determine the distributions of...

- $Y = X + 3$
- $Y = 2X + 3$
- $Y = 3X + 2$

Affine Transformations of Normal (Univariate)

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$ with $a \neq 0$.

If we define $Y = aX + b$, then $Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.

Suppose that $X \sim \mathrm{N}(1, 2)$. Determine the distributions of...

- $Y = X + 3 \quad \Longrightarrow \quad Y \sim \mathrm{N}(1(1) + 3, 1^2(2)) \equiv \mathrm{N}(4, 2)$
- $Y = 2X + 3 \quad \Longrightarrow \quad Y \sim \mathrm{N}(2(1) + 3, 2^2(2)) \equiv \mathrm{N}(5, 8)$
- $Y = 3X + 2 \quad \Longrightarrow \quad Y \sim \mathrm{N}(3(1) + 2, 3^2(2)) \equiv \mathrm{N}(5, 18)$

Normal Density Function (Bivariate)

Given two variables $x, y \in \mathbb{R}$, the bivariate normal pdf is

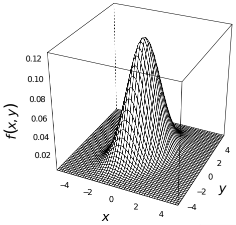
$$f(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \quad (5)$$

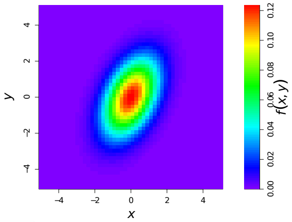
where

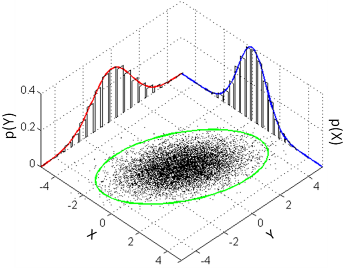
- $\mu_x \in \mathbb{R}$ and $\mu_y \in \mathbb{R}$ are the marginal means
- $\sigma_x \in \mathbb{R}^+$ and $\sigma_y \in \mathbb{R}^+$ are the marginal standard deviations
- $0 \leq |\rho| \leq 1$ is the correlation coefficient

X and Y are marginally normal: $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathrm{N}(\mu_y, \sigma_y^2)$

Example: $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 2$, $\rho = 0.6/\sqrt{2}$

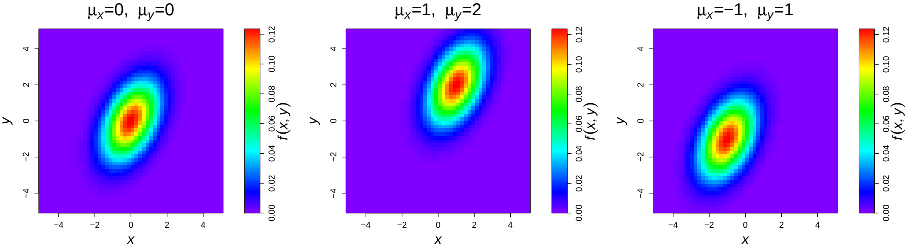






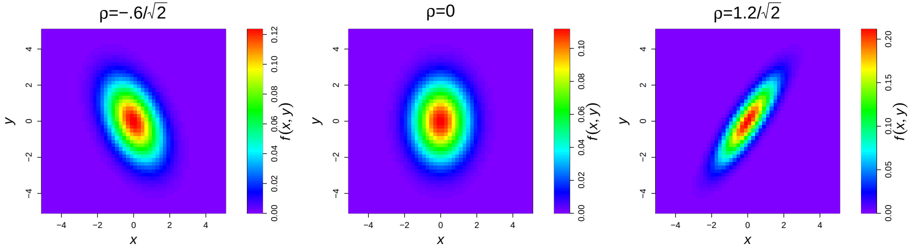
http://en.wikipedia.org/wiki/File:MultivariateNormal.png

Example: Different Means



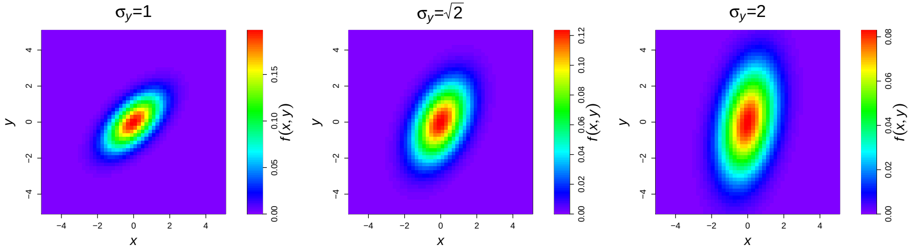
Note: for all three plots $\sigma_x^2 = 1$, $\sigma_y^2 = 2$, and $\rho = 0.6/\sqrt{2}$.

Example: Different Correlations



Note: for all three plots $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, and $\sigma_y^2 = 2$.

Example: Different Variances



Note: for all three plots $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, and $\rho = 0.6/(\sigma_x \sigma_y)$.

Probabilities and Multiple Integration

Probabilities still relate to the area under the pdf:

$$P([a_x \leq X \leq b_x] \text{ and } [a_y \leq Y \leq b_y]) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x,y) \mathrm{d}y \mathrm{d}x \quad (6)$$

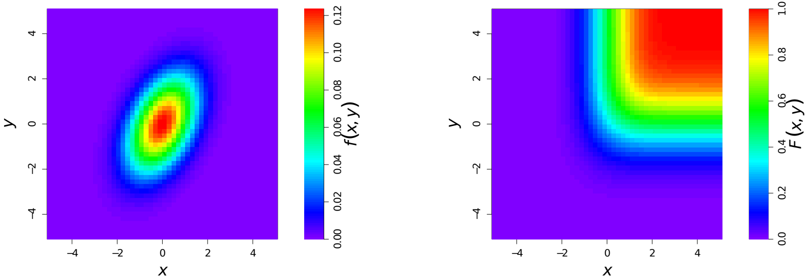
where $\int \int f(x,y)\mathrm{d}y\mathrm{d}x$ denotes the multiple integral of the pdf $f(x,y)$.

Defining $\mathbf{z} = (x,y)$, we can still define the cdf:

$$\begin{aligned} F(\mathbf{z}) &= P(X \leq x \text{ and } Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v) \mathrm{d}v \mathrm{d}u \end{aligned} \quad (7)$$

Normal Distribution Functions (Bivarite)

Helpful figures of bivariate normal pdf and cdf:



Note: $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 2$, and $\rho = 0.6/\sqrt{2}$

Note that the cdf has still has an ogive shape (now in two-dimensions).

Affine Transformations of Normal (Bivariate)

Given $\mathbf{z} = (x, y)'$, suppose that $\mathbf{z} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\boldsymbol{\mu} = (\mu_x, \mu_y)'$ is the 2×1 mean vector
- $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$ is the 2×2 covariance matrix

Let $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ with $\mathbf{A} \neq \mathbf{0}_{2\times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

If we define $\mathbf{w} = \mathbf{A}\mathbf{z} + \mathbf{b}$, then $\mathbf{w} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Conditional Normal (Bivariate)

Given variables X and Y, the conditional distribution of Y given X is

$$f(y|x) = \frac{f(x,y)}{f(x)} \tag{8}$$

where

- $f(x,y)$ is the joint pdf of X and Y
- $f(x)$ is the marginal pdf of X

In the bivariate normal case, we have that

$$Y|X \sim \mathrm{N}(\mu_*, \sigma_*^2) \tag{9}$$

where $\mu_* = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$ and $\sigma_*^2 = \sigma_y^2(1 - \rho^2)$

Derivation of Conditional Normal

To prove Equation (9), simply write out the definition and simplify:

$$
\begin{aligned}
f(y|x) &= \frac{f(x,y)}{f(x)} \\
&= \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\} \Big/ \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)}{\exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\} \Big/ \left(\sigma_X\sqrt{2\pi}\right)} \\
&= \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right] + \frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[\rho^2\frac{\sigma_Y^2}{\sigma_X^2}(x-\mu_X)^2 + (y-\mu_Y)^2 - 2\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)(y-\mu_Y)\right]\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}
\end{aligned}
$$

which completes the proof.

Example #1

A statistics class takes two exams X (Exam 1) and Y (Exam 2) where the scores follow a bivariate normal distribution with parameters:

- $\mu_x = 70$ and $\mu_y = 60$ are the marginal means
- $\sigma_x = 10$ and $\sigma_y = 15$ are the marginal standard deviations
- $\rho = 0.6$ is the correlation coefficient

Suppose we select a student at random. What is the probability that...

(a) the student scores over 75 on Exam 2?

(b) the student scores over 75 on Exam 2, given that the student scored $x = 80$ on Exam 1?

(e) the sum of his/her Exam 1 and Exam 2 scores is over 150?

(f) the student did better on Exam 1 than Exam 2?

(i) $P(5X - 4Y > 150)$?

Example #1 (continued, 1a)

Answer for 1(a):
Note that $Y \sim \mathrm{N}(60, 15^2)$, so the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75) &= P\left(Z > \frac{75 - 60}{15}\right) \\ &= P(Z > 1) \\ &= 1 - P(Z < 1) \\ &= 1 - \Phi(1) \\ &= 1 - 0.8413447 \\ &= 0.1586553 \end{aligned}$$

where $\Phi(x) = \int_{-\infty}^{x} f(z)dz$ with $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denoting the standard normal pdf.

Example #1 (continued, 1b)

Answer for 1(b):

Note that $(Y|X = 80) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 60 + (0.6)(15/10)(80 - 70) = 69$

$\sigma_*^2 = \sigma_Y^2(1 - \rho^2) = 15^2(1 - 0.6^2) = 144$

If a student scored $x = 80$ on Exam 1, the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75|X = 80) &= P\left(Z > \frac{75 - 69}{12}\right) \\ &= P(Z > 0.5) \\ &= 1 - \Phi(0.5) \\ &= 1 - 0.6914625 \\ &= 0.3085375 \end{aligned}$$

Example #1 (continued, 1e)

Answer for 1(e):

Note that $(X+Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_X + \mu_Y = 70 + 60 = 130$

$\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 + 2(0.6)(10)(15) = 505$

The probability that the sum of Exam 1 and Exam 2 is above 150 is

$$\begin{aligned} P(X+Y > 150) &= P\left(Z > \frac{150-130}{\sqrt{505}}\right) \\ &= P\left(Z > 0.8899883\right) \\ &= 1 - \Phi(0.8899883) \\ &= 1 - 0.8132639 \\ &= 0.1867361 \end{aligned}$$

Example #1 (continued, 1f)

Answer for 1(f):

Note that $(X - Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_X - \mu_Y = 70 - 60 = 10$

$\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 - 2(0.6)(10)(15) = 145$

The probability that the student did better on Exam 1 than Exam 2 is

$$\begin{aligned} P(X > Y) &= P(X - Y > 0) \\ &= P\left(Z > \frac{0 - 10}{\sqrt{145}}\right) \\ &= P\left(Z > -0.8304548\right) \\ &= 1 - \Phi(-0.8304548) \\ &= 1 - 0.2031408 \\ &= 0.7968592 \end{aligned}$$

Example #1 (continued, 1i)

Answer for 1(i):

Note that $(5X - 4Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where
$\mu_* = 5\mu_X - 4\mu_Y = 5(70) - 4(60) = 110$
$\sigma_*^2 = 5^2\sigma_X^2 + (-4)^2\sigma_Y^2 + 2(5)(-4)\rho\sigma_X\sigma_Y =$
$25(10^2) + 16(15^2) - 2(20)(0.6)(10)(15) = 2500$

Thus, the needed probability can be obtained using

$$\begin{aligned} P(5X - 4Y > 150) &= P\left(Z > \frac{150 - 110}{\sqrt{2500}}\right) \\ &= P(Z > 0.8) \\ &= 1 - \Phi(0.8) \\ &= 1 - 0.7881446 \\ &= 0.2118554 \end{aligned}$$

Vectors and Matrices

A *vector* is a one-dimensional array: $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}_{n\times 1}$

A *matrix* is a two-dimensional array: $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}_{n\times p}$

The *order* of a matrix refers the to number of rows and columns:

- $\mathbf{a}$ has order n-by-1
- $\mathbf{A}$ has order n-by-p

Matrix Transpose: Definition

We will denote the *transpose* with a prime symbol (i.e., $'$).

The *transpose* of a vector turns a column vector into a row vector:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}_{n\times 1} \iff \mathbf{a}' = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}_{1\times n}$$

The *transpose* of a matrix exchanges rows and columns, such as

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}_{n\times p} \iff \mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{pmatrix}_{p\times n}$$

Matrix Transpose: Example

The transpose of $\mathbf{a} = \begin{pmatrix} 1 \\ 7 \\ 5 \\ 9 \end{pmatrix}_{4\times 1}$ is given by $\mathbf{a}' = \begin{pmatrix} 1 & 7 & 5 & 9 \end{pmatrix}_{1\times 4}$

The transpose of $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 7 & 2 \\ 5 & 7 \\ 9 & 4 \end{pmatrix}_{4\times 2}$ is given by $\mathbf{A}' = \begin{pmatrix} 1 & 7 & 5 & 9 \\ 3 & 2 & 7 & 4 \end{pmatrix}_{2\times 4}$

Matrix Transpose: Properties

Some useful properties of matrix transposes include:

- $(\mathbf{A}')' = \mathbf{A}$
- $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$ (where $\mathbf{A} + \mathbf{B}$ is matrix addition, later defined)
- $(b\mathbf{A})' = b\mathbf{A}'$ (where $b\mathbf{A}$ is scalar multiplication, later defined)
- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ (where $\mathbf{AB}$ is matrix multiplication, later defined)
- $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$ (where $\mathbf{A}^{-1}$ is matrix inverse, later defined)

Matrix Trace: Definition

The *trace* of a square matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}_{p \times p}$ is

$$\mathrm{tr}(\mathbf{A}) = \sum_{j=1}^{p} a_{jj} \tag{10}$$

which is the sum of the diagonal elements.

Matrix Trace: Example

The trace of the matrix $\mathbf{A} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \\ 5 & 9 & 4 & 3 \end{pmatrix}$ is

$$\begin{aligned} \mathrm{tr}(\mathbf{A}) &= 1 + 8 + 6 + 3 \\ &= 18 \end{aligned}$$

Matrix Trace: Properties

Some useful properties of matrix traces include:

- $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}')$
- $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
- $\mathrm{tr}(b\mathbf{A}) = b\mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ if both products are defined
- If $\mathbf{A}$ is symmetric, $\mathrm{tr}(\mathbf{A}) = \sum_{j=1}^{p} \lambda_j$ where λ_j is j-th eigenvalue of $\mathbf{A}$.

Symmetric Matrix: Definition

A *symmetric* matrix is square and symmetric along the main diagonal:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}_{n \times n} \tag{11}$$

with $a_{ij} = a_{ji}$ for all $i \neq j$.

Note that $\mathbf{A} = \mathbf{A}'$ for all symmetric matrices (by definition).

Symmetric Matrix: Example

The matrix $\mathbf{A} = \begin{pmatrix} 9 & 1 & 0 & 4 \\ 1 & 4 & 2 & 1 \\ 0 & 2 & 5 & 6 \\ 4 & 1 & 6 & 8 \end{pmatrix}$ is a symmetric 4×4 matrix.

The matrix $\mathbf{A} = \begin{pmatrix} 9 & 1 & 0 & {\color{red}4} \\ 1 & 4 & 2 & 1 \\ 0 & 2 & 5 & 6 \\ {\color{red}3} & 1 & 6 & 8 \end{pmatrix}$ is NOT a symmetric 4×4 matrix.

Diagonal Matrix

A *diagonal* matrix is a square matrix that has zeros in the off-diagonals:

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_p \end{pmatrix}_{p \times p} \tag{12}$$

We often write $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$ to define a diagonal matrix.

Identity Matrix

The *identity matrix* of order p is a $p \times p$ matrix that has ones along the main diagonal and zeros in the off-diagonals:

$$\mathbf{I}_p = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{p \times p} \tag{13}$$

Note that $\mathbf{I}_p$ is a special type of diagonal matrix.

Zero and One Matrices

A vector or matrix of all zeros will be denoted using the notation:

$$\mathbf{0}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{n \times 1} \qquad \mathbf{0}_{n \times p} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{n \times p}$$

A vector or matrix of all ones will be denoted using the notation:

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1} \qquad \mathbf{1}_{n \times p} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{n \times p}$$

Matrix Equality

Given two matrices of the same order $\mathbf{A} = \{a_{ij}\}_{n\times p}$ and $\mathbf{B} = \{b_{ij}\}_{n\times p}$, we say that $\mathbf{A}$ is equal to $\mathbf{B}$ (written $\mathbf{A} = \mathbf{B}$) if and only if $a_{ij} = b_{ij}\ \forall i, j$.

If $\mathbf{A} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \end{pmatrix}$, then $\mathbf{A} = \mathbf{B}$.

If $\mathbf{A} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 0 \end{pmatrix}$, then $\mathbf{A} \neq \mathbf{B}$.

Matrix Addition and Subtraction: Definition

Given two matrices of the same order $\mathbf{A} = \{a_{ij}\}_{n \times p}$ and $\mathbf{B} = \{b_{ij}\}_{n \times p}$, the addition $\mathbf{A} + \mathbf{B}$ produces $\mathbf{C} = \{c_{ij}\}_{n \times p}$ such that $c_{ij} = a_{ij} + b_{ij}$.

Given two matrices of the same order $\mathbf{A} = \{a_{ij}\}_{n \times p}$ and $\mathbf{B} = \{b_{ij}\}_{n \times p}$, the subtraction $\mathbf{A} - \mathbf{B}$ produces $\mathbf{C} = \{c_{ij}\}_{n \times p}$ such that $c_{ij} = a_{ij} - b_{ij}$.

Note: matrix addition and subtraction is only defined for two matrices of the same order.

Matrix Addition and Subtraction: Example

Given $\mathbf{A} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 5 & 6 & 1 & 7 \\ 1 & 3 & 0 & 2 \\ 2 & 5 & 3 & 5 \end{pmatrix}$, we have that

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 1+5 & 4+6 & 8+1 & 13+7 \\ 2+1 & 8+3 & 11+0 & 2+2 \\ 7+2 & 2+5 & 6+3 & 9+5 \end{pmatrix} = \begin{pmatrix} 6 & 10 & 9 & 20 \\ 3 & 11 & 11 & 4 \\ 9 & 7 & 9 & 14 \end{pmatrix}$$

$$\mathbf{A} - \mathbf{B} = \begin{pmatrix} 1-5 & 4-6 & 8-1 & 13-7 \\ 2-1 & 8-3 & 11-0 & 2-2 \\ 7-2 & 2-5 & 6-3 & 9-5 \end{pmatrix} = \begin{pmatrix} -4 & -2 & 7 & 6 \\ 1 & 5 & 11 & 0 \\ 5 & -3 & 3 & 4 \end{pmatrix}$$

Matrix-Scalar Products: Definition

The *matrix-scalar product* of $\mathbf{A} = \{a_{ij}\}_{n\times p}$ and $b \in \mathbb{R}$ is

$$\mathbf{A}b = b\mathbf{A} = \begin{pmatrix} ba_{11} & ba_{12} & \cdots & ba_{1p} \\ ba_{21} & ba_{22} & \cdots & ba_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ ba_{n1} & ba_{n2} & \cdots & ba_{np} \end{pmatrix}_{n\times p} \tag{14}$$

which is the matrix $\mathbf{C} = \{c_{ij}\}_{n\times p}$ such that $c_{ij} = ba_{ij}$.

Matrix-Scalar Products: Example

Given $\mathbf{A} = \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \end{pmatrix}$ and $b = 2$, we have that

$$\begin{aligned} b\mathbf{A} &= \begin{pmatrix} 1 & 4 & 8 & 13 \\ 2 & 8 & 11 & 2 \\ 7 & 2 & 6 & 9 \end{pmatrix} 2 \\ &= \begin{pmatrix} 2 & 8 & 16 & 26 \\ 4 & 16 & 22 & 4 \\ 14 & 4 & 12 & 18 \end{pmatrix} \end{aligned}$$

Vector Inner Products: Definition

The *inner product* of $\mathbf{x} = (x_1, \ldots, x_n)'$ and $\mathbf{y} = (y_1, \ldots, y_n)'$ is

$$\begin{aligned} \mathbf{x}'\mathbf{y} &= \begin{pmatrix} x_1 \cdots x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \left(\sum_{i=1}^{n} x_i y_i \right)_{1\times 1} \end{aligned} \tag{15}$$

Note that $\mathbf{x}$ and $\mathbf{y}$ must have the same length (i.e., n).

Vector Inner Products: Example

Given $\mathbf{x} = (3, 9, -2, 5)'$ and $\mathbf{y} = (2, 0, 2, 1)'$, we have that

$$\begin{aligned}
\mathbf{x}'\mathbf{y} &= \begin{pmatrix} 3 & 9 & -2 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 2 \\ 1 \end{pmatrix} \\
&= 3(2) + 9(0) - 2(2) + 5(1) \\
&= 7
\end{aligned}$$

Vector Outer Products: Definition

The *outer product* of $\mathbf{x} = (x_1, \ldots, x_m)'$ and $\mathbf{y} = (y_1, \ldots, y_n)'$ is

$$\begin{aligned} \mathbf{x}\mathbf{y}' &= \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \begin{pmatrix} y_1 \cdots y_n \end{pmatrix} \\ &= \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{pmatrix}_{m \times n} \end{aligned} \tag{16}$$

Note that $\mathbf{x}$ and $\mathbf{y}$ can have different lengths (i.e., m and n).

Vector Outer Products: Example

Given $\mathbf{x} = (3, 9, -2, 5)'$ and $\mathbf{y} = (2, 0, 2, 1)'$, we have that

$$\mathbf{x}\mathbf{y}' = \begin{pmatrix} 3 \\ 9 \\ -2 \\ 5 \end{pmatrix} \begin{pmatrix} 2 & 0 & 2 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 6 & 0 & 6 & 3 \\ 18 & 0 & 18 & 9 \\ -4 & 0 & -4 & -2 \\ 10 & 0 & 10 & 5 \end{pmatrix}$$

Matrix-Vector Products: Definition

The *matrix-vector product* of $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ is

$$\begin{aligned} \mathbf{Ax} &= \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^{p} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{p} a_{nj}x_j \end{pmatrix}_{n\times 1} \end{aligned} \tag{17}$$

Note that length of **x** must match number of columns of **A** (i.e., p).

Matrix-Vector Products: Example

Given $\mathbf{A} = \begin{pmatrix} 3 & 4 & 1 \\ 4 & 7 & 5 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix}$, we have that

$$
\begin{aligned}
\mathbf{Ax} &= \begin{pmatrix} 3 & 4 & 1 \\ 4 & 7 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 3(1) + 4(6) + 1(3) \\ 4(1) + 7(6) + 5(3) \end{pmatrix} \\
&= \begin{pmatrix} 30 \\ 61 \end{pmatrix}
\end{aligned}
$$

Matrix-Matrix Products: Definition

The *matrix-matrix product* of $\mathbf{A} = \{a_{ij}\}_{m \times n}$ and $\mathbf{B} = \{b_{jk}\}_{n \times p}$ is

$$\begin{aligned}
\mathbf{AB} &= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{np} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^{n} a_{1j} b_{j1} & \sum_{j=1}^{n} a_{1j} b_{j2} & \cdots & \sum_{j=1}^{n} a_{1j} b_{jp} \\ \sum_{j=1}^{n} a_{2j} b_{j1} & \sum_{j=1}^{n} a_{2j} b_{j2} & \cdots & \sum_{j=1}^{n} a_{2j} b_{jp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} a_{mj} b_{j1} & \sum_{j=1}^{n} a_{mj} b_{j2} & \cdots & \sum_{j=1}^{n} a_{mj} b_{jp} \end{pmatrix}_{m \times p}
\end{aligned} \tag{18}$$

Note that # of rows of **B** must match # of columns of **A** (i.e., n), and note that $\mathbf{AB} \neq \mathbf{BA}$ even if both products are defined.

Matrix-Matrix Products: Example

Given $\mathbf{A} = \begin{pmatrix} 3 & 4 & 1 \\ 4 & 7 & 5 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 6 & 1 \\ 3 & 4 \end{pmatrix}$, we have that

$$
\begin{aligned}
\mathbf{AB} &= \begin{pmatrix} 3 & 4 & 1 \\ 4 & 7 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 6 & 1 \\ 3 & 4 \end{pmatrix} \\
&= \begin{pmatrix} 3(1) + 4(6) + 1(3) & 3(2) + 4(1) + 1(4) \\ 4(1) + 7(6) + 5(3) & 4(2) + 7(1) + 5(4) \end{pmatrix} \\
&= \begin{pmatrix} 30 & 14 \\ 61 & 35 \end{pmatrix}
\end{aligned}
$$

Multiplying by Identity Matrix

Given $\mathbf{A} = \{a_{ij}\}_{m \times n}$, pre-multiplying by the identity matrix returns $\mathbf{A}$

$$\mathbf{I}_m\mathbf{A} = \mathbf{A}$$

and post-multiplying by the identity matrix returns $\mathbf{A}$

$$\mathbf{A}\mathbf{I}_n = \mathbf{A}$$

This is the reason we call $\mathbf{I}_m$ and $\mathbf{I}_n$ “identity” matrices.

Quadratic Forms

The *quadratic form* of a symmetric matrix $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ is

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &= \left(\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij} \right)_{1\times 1} \end{aligned} \tag{19}$$

where $\mathbf{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}'$ is any arbitrary vector of length n.

Positive, Negative, and Semi-Definite Matrices

A symmetric matrix $\mathbf{A} = \{a_{ij}\}_{n \times n}$ is said to be

- *positive definite* if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for every $\mathbf{x} \neq \mathbf{0}_n$
- *positive semi-definite* if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for every $\mathbf{x} \neq \mathbf{0}_n$
- *negative definite* if $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ for every $\mathbf{x} \neq \mathbf{0}_n$
- *negative semi-definite* if $\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$ for every $\mathbf{x} \neq \mathbf{0}_n$

Note if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for some $\mathbf{x}$ and $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ for other $\mathbf{x}$, then $\mathbf{A}$ is said to be an *indefinite* matrix.

Matrix Definiteness: Example

The matrix $\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ is positive definite:

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 \end{pmatrix} \\ &= 2x_1^2 - 2x_1x_2 + 2x_2^2 \\ &= x_1^2 + x_2^2 + (x_1 - x_2)^2 \\ &\geq 0 \end{aligned}$$

with the equality holding only when $x_1 = x_2 = 0$.

Matrix Definiteness: Properties

Let λ_j denote the j-th eigenvalue of $\mathbf{A}$ for $j \in \{1, \ldots, n\}$.

Some useful properties of matrix definiteness include:

- If $\mathbf{A}$ is positive definite, then $\lambda_j > 0 \ \forall j$
- If $\mathbf{A}$ is positive semi-definite, then $\lambda_j \geq 0 \ \forall j$
- If $\mathbf{A}$ is negative definite, then $\lambda_j < 0 \ \forall j$
- If $\mathbf{A}$ is negative semi-definite, then $\lambda_j \leq 0 \ \forall j$
- If $\mathbf{A}$ is indefinite, then $\lambda_i > 0$ and $\lambda_j < 0$ for some $i \neq j$

Overview of Matrix Decompositions

A *matrix decomposition* decomposes (i.e., separates) a given matrix into a matrix multiplication of two (or more) simpler matrices.

Matrix decompositions are useful for many things:

- Solving systems of equations
- Obtaining low-rank approximations
- Finding important features of data

We will briefly discuss four matrix decompositions:

- Eigenvalue Decomposition
- Cholesky Decomposition
- Singular Value Decomposition
- QR Decomposition

Eigenvalue Decomposition

The *eigenvalue decomposition* (EVD) decomposes a symmetric matrix $\mathbf{A} = \{a_{ij}\}_{n \times n}$ into a product of three matrices:

$$\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}' \tag{20}$$

such that

- $\boldsymbol{\Gamma} = (\gamma_1 \cdots \gamma_n)_{n \times n}$ where $\gamma_j = (\gamma_{1j}, \ldots, \gamma_{nj})'$ is j-th eigenvector
- $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ where λ_j is j-th eigenvalue

Note that $\boldsymbol{\Gamma}$ is an orthogonal matrix: $\boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \mathbf{I}_n$

Cholesky Decomposition

The *Cholesky decomposition* (CD) decomposes a positive definite matrix $\mathbf{A} = \{a_{ij}\}_{n \times n}$ into a product of a two matrices:

$$\mathbf{A} = \mathbf{L}\mathbf{L}' \tag{21}$$

where

- $\mathbf{L} = \begin{pmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nn} \end{pmatrix}$ is a lower (left) triangular matrix

Singular Value Decomposition

The *singular value decomposition* (SVD) decomposes any matrix $\mathbf{A} = \{a_{ij}\}_{n \times p}$ into a product of three matrices:

$$\mathbf{A} = \mathbf{USV}' \tag{22}$$

such that

- $\mathbf{U} = (\mathbf{u}_1 \cdots \mathbf{u}_r)_{n \times r}$ where $\mathbf{u}_k = \{u_{ik}\}_{n \times 1}$ is k-th left singular vector
- $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_r)$ where $s_k > 0$ is k-th singular value
- $\mathbf{V} = (\mathbf{v}_1 \cdots \mathbf{v}_r)_{p \times r}$ where $\mathbf{v}_k = \{v_{jk}\}_{p \times 1}$ is k-th right singular vector
- $r \leq \min(m, n)$ and $r = \min(m, n)$ if $\mathbf{A}$ is full-rank

Note that $\mathbf{U}$ and $\mathbf{V}$ are columnwise orthogonal: $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$

QR Decomposition

The *QR decomposition* (QRD) decomposes any long (i.e., $n \geq p$) matrix $\mathbf{A} = \{a_{ij}\}_{n \times p}$ into a product of two matrices:

$$\begin{aligned}\mathbf{A} &= \mathbf{QR} \\ &= \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix} \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{0}_{(n-p) \times p} \end{pmatrix} \\ &= \mathbf{Q}_1 \mathbf{R}_1 \end{aligned} \tag{23}$$

such that

- $(\mathbf{Q}_1)_{n \times p}$ and $(\mathbf{Q}_2)_{n \times (n-p)}$ are columnwise orthogonal
- $\mathbf{R}_1 = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ 0 & r_{22} & r_{23} & \cdots & r_{2p} \\ 0 & 0 & r_{33} & \cdots & r_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{pp} \end{pmatrix}$ is upper (right) triangular matrix

Matrix Determinant: Definition

The determinant of a square matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$ is a real-valued function from $\mathbb{R}^{p \times p} \to \mathbb{R}$, and is typically denoted by $|\mathbf{A}|$ or $\det(\mathbf{A})$.

Determinants provide information about systems of linear equations:

- Suppose that $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{x} \in \mathbb{R}^{p \times 1}$, and $\mathbf{b} \in \mathbb{R}^{p \times 1}$
- System $\mathbf{Ax} = \mathbf{b}$ has a unique solution if and only if $|\mathbf{A}| \neq 0$

Determinants provide information about linear transformations:

- Magnitude of $|\mathbf{A}|$ is the transformation's *scale factor*
- Sign of $|\mathbf{A}|$ is the transformation's *orientation*

Matrix Determinant: Calculation

- For 1×1 matrix $\mathbf{A} = (a)$, we have
  $|\mathbf{A}| = a$

- For 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we have
  $|\mathbf{A}| = ad - bc$

- For 3×3 matrix $\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, we have
  $|\mathbf{A}| = aei + bfg + cdh - (ceg + bdi + afh)$

Matrix Determinant: Calculation (continued)

For $p \times p$ matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}$, we have

$|\mathbf{A}| = \sum_{j=1}^{p} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{p} (-1)^{i+j} a_{ij} M_{ij}$

where

- $M_{ij} = |\mathbf{A}_{-ij}|$ is the *minor* corresponding to cell (i, j) of $\mathbf{A}$
- $(-1)^{i+j} M_{ij}$ is the *cofactor* corresponding to cell (i, j) of $\mathbf{A}$
- $\mathbf{A}_{-ij}$ is the $(p-1) \times (p-1)$ matrix formed by deleting the i-th row and j-th column of $\mathbf{A}$

Note: can use any column (or row) to define the determinant of $\mathbf{A}$.

Properties of Matrix Determinants

Some useful properties of matrix determinants include:

- $|\mathbf{A}| = |\mathbf{A}'|$
- $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$ (where $\mathbf{A}^{-1}$ is defined on the next slide)
- $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$ (if $\mathbf{A}$ and $\mathbf{B}$ are both square)
- $|b\mathbf{A}| = b^p|\mathbf{A}|$ (if $b \in \mathbb{R}$ and $\mathbf{A}$ is $p \times p$)
- If $\mathbf{A}$ is symmetric, $|\mathbf{A}| = \prod_{j=1}^{p} \lambda_j$ where λ_j is j-th eigenvalue of $\mathbf{A}$.

Matrix Inverses: Definition

A square (not necessarily symmetric) matrix $\mathbf{A} = \{a_{ij}\}_{n \times n}$ is *invertible* (or *nonsingular*) if there exists another matrix $\mathbf{B} = \{b_{ij}\}_{n \times n}$ such that

$$\mathbf{AB} = \mathbf{I}_n \tag{24}$$

where $\mathbf{I}_n$ is the $n \times n$ *identity matrix*.

If $\mathbf{B}$ exists, the matrix $\mathbf{B}$ is called the *inverse* of the matrix $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$ (so that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$).

Matrix Inverses: Example

Given $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$, the inverse is $\mathbf{A}^{-1} = \begin{pmatrix} -1/5 & 3/5 \\ 2/5 & -1/5 \end{pmatrix}$:

$$\mathbf{A}\mathbf{A}^{-1} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -1/5 & 3/5 \\ 2/5 & -1/5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{A}^{-1}\mathbf{A} = \begin{pmatrix} -1/5 & 3/5 \\ 2/5 & -1/5 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Matrix Inverses: Properties

Some useful properties of matrix inverses include:

- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(b\mathbf{A})^{-1} = b^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$
- $\mathbf{A}^{-1} = \mathbf{A}'$ if and only if $\mathbf{A}$ is orthogonal
- $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ if both $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist
- $\mathbf{A}^{-1}$ exists only if $|\mathbf{A}| \neq 0$
- If $\mathbf{A}$ is positive definite, then $\mathbf{A}^{-1} = \mathbf{\Gamma\Lambda}^{-1}\mathbf{\Gamma}' = (\mathbf{L}^{-1})'\mathbf{L}^{-1}$, where $\mathbf{\Gamma\Lambda\Gamma}'$ and $\mathbf{LL}'$ denote the EVD and CD of $\mathbf{A}$, respectively

Normal Density Function (Multivariate)

Given $\mathbf{x} = (x_1, \ldots, x_p)$ with $x_j \in \mathbb{R}\ \forall j$, the multivariate normal pdf is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \quad (25)$$

where

- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ is the $p \times 1$ mean vector
- $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$ is the $p \times p$ covariance matrix

Write $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or $\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote $\mathbf{x}$ is multivariate normal.

Some Multivariate Normal Properties

The mean and covariance parameters have the following restrictions:

- $\mu_j \in \mathbb{R}$ for all j
- $\sigma_{jj} > 0$ for all j
- $\sigma_{ij}^2 \leq \sigma_{ii}\sigma_{jj}$ for any $i, j \in \{1, \ldots, p\}$ (Cauchy-Schwarz)

$\boldsymbol{\Sigma}$ is assumed to be positive definite so that $\boldsymbol{\Sigma}^{-1}$ exists.

Marginals are normal: $x_j \sim \mathrm{N}(\mu_j, \sigma_{jj})$ for all $j \in \{1, \ldots, p\}$.

Multivariate Normal Probabilities

Probabilities still relate to the area under the pdf:

$$P(a_j \leq X_j \leq b_j \; \forall j) = \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(\mathbf{x}) \mathrm{d}x_p \cdots \mathrm{d}x_1 \tag{26}$$

where $\int \cdots \int f(\mathbf{x}) \mathrm{d}x_p \cdots \mathrm{d}x_1$ denotes the multiple integral $f(\mathbf{x})$.

We can still define the cdf of $\mathbf{x} = (x_1, \ldots, x_p)$:

$$\begin{aligned} F(\mathbf{x}) &= P(X_j \leq x_j \; \forall j) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(\mathbf{u}) \mathrm{d}u_p \cdots \mathrm{d}u_1 \end{aligned} \tag{27}$$

Affine Transformations of Normal (Multivariate)

Suppose that $\mathbf{x} = (x_1, \ldots, x_p)'$ and that $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\boldsymbol{\mu} = \{\mu_j\}_{p\times 1}$ is the mean vector
- $\boldsymbol{\Sigma} = \{\sigma_{ij}\}_{p\times p}$ is the covariance matrix

Let $\mathbf{A} = \{a_{ij}\}_{n\times p}$ and $\mathbf{b} = \{b_i\}_{n\times 1}$ with $\mathbf{A} \neq \mathbf{0}_{n\times p}$.

If we define $\mathbf{w} = \mathbf{A}\mathbf{x} + \mathbf{b}$, then $\mathbf{w} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Note: linear combinations of normal variables are normally distributed.

Multivariate Conditional Distributions

Given variables $\mathbf{x} = (x_1, \ldots, x_p)'$ and $\mathbf{y} = (y_1, \ldots, y_q)'$, we have

$$f(\mathbf{y}|\mathbf{x}) = \frac{f(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})} \tag{28}$$

where

- $f(\mathbf{y}|\mathbf{x})$ the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ is
- $f(\mathbf{x}, \mathbf{y})$ is the joint pdf of $\mathbf{x}$ and $\mathbf{y}$
- $f(\mathbf{x})$ is the marginal pdf of $\mathbf{x}$

Conditional Normal (Multivariate)

Suppose that $\mathbf{z} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\mathbf{z} = (\mathbf{x}', \mathbf{y}')' = (x_1, \ldots, x_p, y_1, \ldots, y_q)'$
- $\boldsymbol{\mu} = (\boldsymbol{\mu}_x', \boldsymbol{\mu}_y')' = (\mu_{1x}, \ldots, \mu_{px}, \mu_{1y}, \ldots, \mu_{qy})'$
  Note: $\boldsymbol{\mu}_x$ is mean vector of $\mathbf{x}$, and $\boldsymbol{\mu}_y$ is mean vector of $\mathbf{y}$
- $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}' & \boldsymbol{\Sigma}_{yy} \end{pmatrix}$ where $(\boldsymbol{\Sigma}_{xx})_{p\times p}$, $(\boldsymbol{\Sigma}_{yy})_{q\times q}$, and $(\boldsymbol{\Sigma}_{xy})_{p\times q}$,
  Note: $\boldsymbol{\Sigma}_{xx}$ is covariance matrix of $\mathbf{x}$, $\boldsymbol{\Sigma}_{yy}$ is covariance matrix of $\mathbf{y}$, and $\boldsymbol{\Sigma}_{xy}$ is covariance matrix of $\mathbf{x}$ and $\mathbf{y}$

In the multivariate normal case, we have that

$$\mathbf{y}|\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*) \tag{29}$$

where $\boldsymbol{\mu}_* = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{xy}'\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$ and $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{xy}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$

Example #4

Each Delicious Candy Company store makes 3 size candy bars: regular (X_1), fun size (X_2), and big size (X_3).

Assume the weight (in ounces) of the candy bars (X_1, X_2, X_3) follow a multivariate normal distribution with parameters:

- $\boldsymbol{\mu} = \begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix}$

Suppose we select a store at random. What is the probability that...

(a) the weight of a regular candy bar is greater than 8 oz?

(b) the weight of a regular candy bar is greater than 8 oz, given that the fun size bar weighs 1 oz and the big size bar weighs 10 oz?

(c) $P(4X_1 - 3X_2 + 5X_3 < 63)$?

Example #4 (continued, 4a)

Answer for 4(a):
Note that $X_1 \sim \mathrm{N}(5, 4)$

So, the probability that the regular bar is more than 8 oz is

$$\begin{aligned} P(X_1 > 8) &= P\left(Z > \frac{8-5}{2}\right) \\ &= P(Z > 1.5) \\ &= 1 - \Phi(1.5) \\ &= 1 - 0.9331928 \\ &= 0.0668072 \end{aligned}$$

Example #4 (continued, 4b)

Answer for 4(b):
$(X_1 | X_2 = 1, X_3 = 10)$ is normally distributed, see Equation (29).

The conditional mean of $(X_1 | X_2 = 1, X_3 = 10)$ is given by

$$\begin{aligned}
\mu_* &= \mu_{X_1} + \boldsymbol{\Sigma}'_{12}\boldsymbol{\Sigma}_{22}^{-1}(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}}) \\
&= 5 + \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 1-3 \\ 10-7 \end{pmatrix} \\
&= 5 + \begin{pmatrix} -1 & 0 \end{pmatrix} \frac{1}{32} \begin{pmatrix} 9 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix} \\
&= 5 + 24/32 \\
&= 5.75
\end{aligned}$$

Example #4 (continued, 4b)

Answer for 4(b) continued:

The conditional variance of $(X_1|X_2 = 1, X_3 = 10)$ is given by

$$\begin{aligned}
\sigma_*^2 &= \sigma_{X_1}^2 - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12} \\
&= 4 - \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 9 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\
&= 4 - \begin{pmatrix} -1 & 0 \end{pmatrix} \frac{1}{32} \begin{pmatrix} 9 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\
&= 4 - 9/32 \\
&= 3.71875
\end{aligned}$$

Example #4 (continued, 4b)

Answer for 4(b) continued:

So, if the fun size bar weighs 1 oz and the big size bar weighs 10 oz, the probability that the regular bar is more than 8 oz is

$$\begin{aligned} P(X_1 > 8 | X_2 = 1, X_3 = 10) &= P\left(Z > \frac{8 - 5.75}{\sqrt{3.71875}} \right) \\ &= P(Z > 1.166767) \\ &= 1 - \Phi(1.166767) \\ &= 1 - 0.8783477 \\ &= 0.1216523 \end{aligned}$$

Example #4 (continued, 4c)

Answer for 4(c):
$(4X_1 - 3X_2 + 5X_3)$ is normally distributed.

The expectation of $(4X_1 - 3X_2 + 5X_3)$ is given by

$$\begin{aligned} \mu_* &= 4\mu_{X_1} - 3\mu_{X_2} + 5\mu_{X_3} \\ &= 4(5) - 3(3) + 5(7) \\ &= 46 \end{aligned}$$

Example #4 (continued, 4c)

Answer for 4(c) continued:

The variance of $(4X_1 - 3X_2 + 5X_3)$ is given by

$$
\begin{aligned}
\sigma_*^2 &= \begin{pmatrix} 4 & -3 & 5 \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} 4 \\ -3 \\ 5 \end{pmatrix} \\
&= \begin{pmatrix} 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix} \begin{pmatrix} 4 \\ -3 \\ 5 \end{pmatrix} \\
&= \begin{pmatrix} 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} 19 \\ -6 \\ 39 \end{pmatrix} \\
&= 289
\end{aligned}
$$

Example #4 (continued, 4c)

Answer for 4(c) continued:

So, the needed probability can be obtained as

$$\begin{aligned} P(4X_1 - 3X_2 + 5X_3 < 63) &= P\left(Z < \frac{63-46}{\sqrt{289}}\right) \\ &= P(Z < 1) \\ &= \Phi(1) \\ &= 0.8413447 \end{aligned}$$