Introduction to Linear Algebra

ECON 5753 — University of Arkansas

Prof. Kyle Butts

January 2025

Matrix Fundamentals

Matrix Transformations

Inverse of a Matrix

Normal Random Variables

Derivatives of matrix expressions

Data

The basic currency of *forecasting* is data. Data typically looks like:

Columns are variables ↓

Rows are observations →

mpg	cyl	hp	wt	am	gear
21.0	6	110	2.620	1	4
21.0	6	110	2.875	1	4
22.8	4	93	2.320	1	4
21.4	6	110	3.215	0	3
18.7	8	175	3.440	0	3

The basic currency of *linear algebra* are matrices. Matrices look like this:

$$\begin{bmatrix} 21.0 & 6 & 110 & 2.620 & 1 & 4 \\ 21.0 & 6 & 110 & 2.875 & 1 & 4 \\ 22.8 & 4 & 93 & 2.320 & 1 & 4 \\ 21.4 & 6 & 110 & 3.215 & 0 & 3 \\ 18.7 & 8 & 175 & 3.440 & 0 & 3 \end{bmatrix}$$

Matrix notations

$$\begin{bmatrix} 21.0 & 6 & 110 & 2.620 & 1 & 4 \\ 21.0 & 6 & 110 & 2.875 & 1 & 4 \\ 22.8 & 4 & 93 & 2.320 & 1 & 4 \\ 21.4 & 6 & 110 & 3.215 & 0 & 3 \\ 18.7 & 8 & 175 & 3.440 & 0 & 3 \end{bmatrix}$$

Let's call this matrix, $\boldsymbol{A}$. Each element of $\boldsymbol{A}$ can be referred to as $A_{i,j}$ where i is the *row* and j is the *column*.

Matrix notations

$$\begin{bmatrix} 21.0 & 6 & 110 & 2.620 & 1 & 4 \\ 21.0 & 6 & 110 & 2.875 & 1 & 4 \\ 22.8 & 4 & 93 & 2.320 & 1 & 4 \\ 21.4 & 6 & 110 & 3.215 & 0 & 3 \\ 18.7 & 8 & 175 & 3.440 & 0 & 3 \end{bmatrix}$$

Let's call this matrix, $\boldsymbol{A}$. Each element of $\boldsymbol{A}$ can be referred to as $A_{i,j}$ where i is the *row* and j is the *column*.

- Unit i's data is written $A_{i,\cdot}$
- Variable j is written $A_{\cdot,j}$

Matrix Fundamentals

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

The Size/Dimension a matrix: (rows) $\times$ (columns). E.g. $\boldsymbol{A}$ is a 2×3 matrix.

- The element in row i and column j is referred to as a_{ij} or A_{ij}.
- A square matrix is one with the same number of rows and columns.

Vectors

A column vector, often just "vector", is a $N \times 1$ matrix:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$$

In our previous discussion, each variable in a dataset can be thought of as a *column vector*. A *matrix*, then, is a collection of variables (vectors) glued together

Matrix Addition and Subtraction

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \text{ and } B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

Dimensions must match:

$$(r \times c) \pm (r \times c) \implies (r \times c)$$

A and B are both 2×3 matrices, so we can add and subtract them:

$$A \pm B = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & a_{13} \pm b_{13} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & a_{23} \pm b_{23} \end{bmatrix}$$

Scalar Multiplication

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

For any scalar c (a number like 2 or -4):

$$cA = \begin{bmatrix} c * a_{11} & c * a_{12} & c * a_{13} \\ c * a_{21} & c * a_{22} & c * a_{23} \end{bmatrix}$$

Scalar Multiplication

This is particularly useful when working with vectors,

$$\beta v = \begin{bmatrix} \beta v_1 \\ \beta v_2 \\ \vdots \\ \beta v_N \end{bmatrix}$$

Regression as matrix operations

The most populate forecasting method we will learn is the *linear regression.* Say we want to predict `mpg` given other variables about the car (`hp` and `wt`).

To do so, we take variables, $A_{\cdot,j}$, and multiply it by their slope parameter to make predictions:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \alpha + \begin{bmatrix} 110 \\ 110 \\ 93 \\ 110 \\ 175 \end{bmatrix} \beta_{\texttt{hp}} + \begin{bmatrix} 2.620 \\ 2.875 \\ 2.320 \\ 3.215 \\ 3.440 \end{bmatrix} \beta_{\texttt{wt}}$$

The most populate forecasting method we will learn is the *linear regression*. Say we want to predict `mpg` given other variables about the car (`hp` and `wt`).

The most populate forecasting method we will learn is the *linear regression.* Say we want to predict `mpg` given other variables about the car (`hp` and `wt`).

To do so, we take variables, $A_{\cdot,j}$, and multiply it by their slope parameter to make predictions:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \alpha + \begin{bmatrix} 110 \\ 110 \\ 93 \\ 110 \\ 175 \end{bmatrix} \beta_{\texttt{hp}} + \begin{bmatrix} 2.620 \\ 2.875 \\ 2.320 \\ 3.215 \\ 3.440 \end{bmatrix} \beta_{\texttt{wt}}$$

Using our rules of scalar multiplication and addition of vectors, we get:

$$\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}\alpha+\begin{bmatrix}110\\110\\93\\110\\175\end{bmatrix}\beta_{\mathtt{hp}}+\begin{bmatrix}2.620\\2.875\\2.320\\3.215\\3.440\end{bmatrix}\beta_{\mathtt{wt}}=\begin{bmatrix}\alpha+110\beta_{\mathtt{hp}}+2.620\beta_{\mathtt{wt}}\\\alpha+110\beta_{\mathtt{hp}}+2.875\beta_{\mathtt{wt}}\\\alpha+93\beta_{\mathtt{hp}}+2.320\beta_{\mathtt{wt}}\\\alpha+110\beta_{\mathtt{hp}}+3.215\beta_{\mathtt{wt}}\\\alpha+175\beta_{\mathtt{hp}}+3.440\beta_{\mathtt{wt}}\end{bmatrix}$$

This is a *linear combination* of the columns of the matrix

Matrix Multiplication

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \text{ and } D = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \end{bmatrix}$$

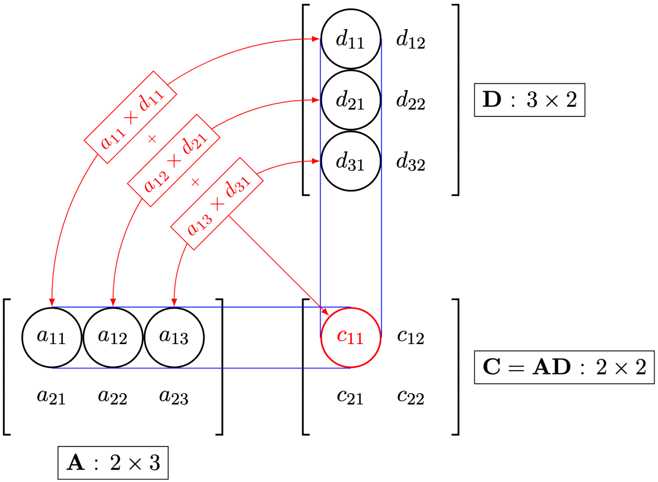
Inner Dimensions must match:

$$(r \times \underline{c}) \times (\underline{c} \times p) \implies (r \times p)$$

A is a 2×3 and D is a 3×2 matrix, so we can multiply (the 2s are equal):

$$A \times D = \begin{bmatrix} a_{11}d_{11} + a_{12}d_{21} + a_{13}d_{31} & a_{11}d_{12} + a_{12}d_{22} + a_{13}d_{32} \\ a_{11}d_{11} + a_{12}d_{21} + a_{13}d_{31} & a_{11}d_{12} + a_{12}d_{22} + a_{13}d_{32} \end{bmatrix}$$

Matrix Multiplication



Matrix Multiplication Practice

$$A = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 3 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 2 \end{bmatrix}$$

What is $A \times B$? What is $B \times A$?

Matrix Multiplication in Regression

It turns out, we can write out our linear combination as a matrix times a vector

$$\begin{bmatrix} 1 & 110 & 2.620 \\ 1 & 110 & 2.875 \\ 1 & 93 & 2.320 \\ 1 & 110 & 3.215 \\ 1 & 175 & 3.440 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_{\texttt{hp}} \\ \beta_{\texttt{wt}} \end{bmatrix}$$

Verify that this product creates the same result

Why call it "multiplication"

There are two rules we learn about regular multiplication:

1. associative: $(ab)c = a(bc)$;
2. distributive: $a(b + c) = ab + ac$; and
3. commutative: $ab = ba$

Why call it "multiplication"

There are two rules we learn about regular multiplication:

1. associative: $(ab)c = a(bc)$;
2. distributive: $a(b + c) = ab + ac$; and
3. commutative: $ab = ba$

Matrices with multiplication and addition are associative, distributive, but *NOT* commutative:

1. $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
2. $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$
3. However, $\boldsymbol{AB} \neq \boldsymbol{AB}$ (in some cases, this holds but not generally)

Commutative

$$\boldsymbol{AB} \neq \boldsymbol{AB}$$

In some cases, one of these two might not even exist. E.g.

- $\boldsymbol{A}$ is a 3×2 matrix and $\boldsymbol{B}$ is a 2×1 matrix.

$\boldsymbol{AB}$ is well-defined (since the number of columns of $\boldsymbol{A} =$ the number of rows of $\boldsymbol{B}$), but $\boldsymbol{BA}$ is not well-defined.

Identity Matrix

In the same way that the number 1 is special in multiplication of numbers (the ‘identity’), the identity matrix takes the following form:

$$\mathbb{I}_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

The identity matrix has the property that $\mathbb{I}_n \boldsymbol{A} = \boldsymbol{A}\mathbb{I}_n A = \boldsymbol{A}$

Uses of Matrix Multiplication

Say I have some variable in my dataset x and I want to know the sum of x (or $1/n*$ the sum to get the sample mean).

Take a few moments and try and think about what the matrix $\boldsymbol{S}$ would need to be to calculate the sample mean of x:

$$\boldsymbol{S} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{n} \left(x_1 + \cdots + x_n \right)$$

$$\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{n}\left(x_1 + \cdots + x_n\right)$$

Answer: The row-vector consisting of all 1s

Uses of Matrix Multiplication

$$\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{n}\left(x_1 + \cdots + x_n\right)$$

Answer: The row-vector consisting of all 1s

The column vector consisting of all 1s is often called “iota”, $\iota = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

The row vector is just ι flipped on it’s side

Transpose

The transpose of a vector will do this “flipping” of vectors and matrices. We denote the transpose as $v^\top$ or v'

- The latter can be confused with ‘derivative’, but is easier to write and arguably standard.

For vectors, the transpose is

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \iff x' = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$$

Transpose

For matrices, the transpose is

$$\boldsymbol{A} = \begin{bmatrix} A_{\cdot,1} & \vdots & A_{\cdot,K} \end{bmatrix} \iff \boldsymbol{A}' = \begin{bmatrix} A'_{\cdot,1} \\ \cdots \\ A'_{\cdot,K} \end{bmatrix}$$

The i-th column of $\boldsymbol{A}$ becomes the i-th row of $\boldsymbol{A}'$

Transpose rules

One rule worth knowing is $(ab)' = b'a'$

- I kind of remember it since transposing “flips” the rows and columns, you also flip the order of the vectors/matrices.

Sample mean

With this, we can write out sample mean more simply as

$$\bar{x} = \frac{1}{n}\iota' x$$

Or, if we wanted the means of multiple varaibles, we could put them in a matrix and

$$\frac{1}{n}\iota' \boldsymbol{X}$$

would be a row-vector of sample means

Sample variance

Recall from your introductory statistics course, the sample variance of a variable is given by

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

Let $\tilde{x}_i \equiv x_i - \bar{x}$. For a minute, think about how you might calculate the sum of squares using the vector $\tilde{x}$

Sample variance

$$s^2 = \frac{1}{n-1}\tilde{x}'\tilde{x} = \frac{1}{n-1}\begin{bmatrix}\tilde{x}_1 & \dots & \tilde{x}_n\end{bmatrix}\begin{bmatrix}\tilde{x}_1 \\ \vdots \\ \tilde{x}_n\end{bmatrix}$$

Similarly, we can write the covariance of two-variables as

$$\frac{1}{n-1}\tilde{x}'\tilde{y} = \frac{1}{n-1}\sum_i \tilde{x}_i\tilde{y}_i$$

Variance Covariance Matrix

Consider a matrix of variable where each column is a (de-meaned) sample.

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ x_2 - \bar{x} & y_2 - \bar{y} & z_2 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} & z_n - \bar{z} \end{bmatrix},$$

where $\bar{x}$ is the mean of variable x.

Variance Covaraince Matrix

The Variance-Covariance Matrix is $\frac{1}{n-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}}$

$$= \frac{1}{n-1}\begin{bmatrix} \sum_{i=1}^n (x_i - \bar{x})^2 & \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) & \sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z}) \\ \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) & \sum_{i=1}^n (y_i - \bar{y})^2 & \sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z}) \\ \sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x}) & \sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y}) & \sum_{i=1}^n (z_i - \bar{z})^2 \end{bmatrix}$$

$$= \begin{bmatrix} Var(x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Var(y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Var(z) \end{bmatrix}$$

Matrix Fundamentals

Matrix Transformations

Inverse of a Matrix

Normal Random Variables

Derivatives of matrix expressions

Matrix Times a Vector (Transformations)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

An $n \times n$ matrix, A, times a $n \times 1$ vector, x, is a transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. So A takes x, rotates it around and/or shrinks or extends the line.

Matrix Times a Vector (Transformations)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

An $n \times n$ matrix, A, times a $n \times 1$ vector, x, is a transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. So A takes x, rotates it around and/or shrinks or extends the line.

In general,

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} \in \mathbb{R}^2$$

Our identity matrix, is a boring transformation:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 * 2 + 0 * 3 \\ 0 * 2 + 1 * 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Transformation Examples

Reflection on the Y-axis:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}$$

Transformation Examples

Reflection 90 degrees clockwise:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix}$$

Enlargement by scale factor a in the x direction and scale factor b in the y direction:

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax \\ by \end{bmatrix}$$

Combination of Transformations

Let's say I want to rotate a vector 90 degrees clockwise and then keep only the x direction (i.e. scale the y by 0.)

- I just multiply the matrices in the order I want to do them:

Let's try:

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{\text{Keep only } x} \overbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}^{\text{Rotate 90 degrees clockwise}} \begin{bmatrix} x \\ y \end{bmatrix} =$$

Let's say I want to rotate a vector 90 degrees clockwise and then keep only the x direction (i.e. scale the y by 0.)

- I just multiply the matrices in the order I want to do them:

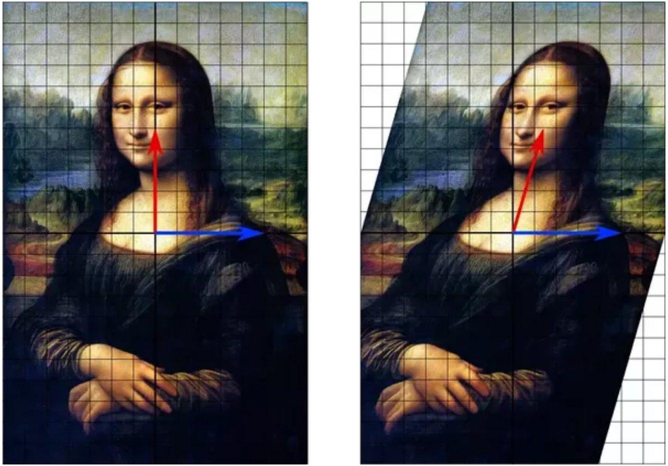
Let's try:

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{\text{Keep only } x} \overbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}^{\text{Rotate 90 degrees clockwise}} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

Determinant of a Matrix

When we rotate and scale an image, we are just doing many many vectors times a transformation matrix.

The determinant asks how much does the area change with our transformation:



Formula for 2×2 Determinant

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

The Determinant of A is given by:

$$\det(A) = a_{11} * a_{22} - a_{12} * a_{21}$$

- If $\det(A) = 1$, then the transformation preserves area
- If $\det(A)$ is greater than/smaller than 1, then the transformation grows/shrinks area.
- If $\det(A) = 0$, then the transformation area shrinks to zero (you lose dimensions).

Matrix Fundamentals

Matrix Transformations

Inverse of a Matrix

Normal Random Variables

Derivatives of matrix expressions

Matrix Inverse

The final necessary linear algebra concept is that of the matrix inverse.

- It is equivalent to $1/x * x = 1$

A *square matrix* (i.e., dimension $n \times n$) $\boldsymbol{A}$ has an inverse $\boldsymbol{A}^{-1}$ if

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \mathbb{I}_n = \boldsymbol{A}\boldsymbol{A}^{-1}$$

Inverse of Matrix

$$\boldsymbol{A}^{-1}\boldsymbol{A}x = \mathbb{I}_n x = x$$

The inverse of a matrix "undoes" the transformation done by A, i.e $AA^{-1}x = A^{-1}Ax = x$.

Inverse of Matrix

$$\boldsymbol{A}^{-1}\boldsymbol{A}x = \mathbb{I}_n x = x$$

The inverse of a matrix "undoes" the transformation done by A, i.e $AA^{-1}x = A^{-1}Ax = x$.

If the determinant of a matrix is 0, then the transformation does not have an inverse.

- For example, the matrix that only keeps the x component can't be inverted (what is the correct y value?)

Inverse of 2x2 Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

For $2x2$ matrices, there is a nice formula for the inverse:

$$\boldsymbol{A}^{-1} = \frac{1}{\underbrace{a_{11}a_{22} - a_{12}a_{21}}_{=\det(\boldsymbol{A})}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Inverse Example

$$\boldsymbol{A} = \begin{bmatrix} 2 & 4 \\ -4 & 10 \end{bmatrix}$$

Find the inverse of $\boldsymbol{A}$ and verify it is indeed the inverse of $\boldsymbol{A}$

Inverse Example

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

What is the determinant of B? Does B have an inverse?

Matrix Fundamentals

Matrix Transformations

Inverse of a Matrix

Normal Random Variables

Derivatives of matrix expressions

Repeated Sampling

When conducting forecasts, we want to express uncertainty around our best guess, what is called statistical inference.

We use the repeated sampling framework to think about this:

- We collect one sample of data and estimate a model
- Imagine collecting *many many* samples in the same way and estimating the model for each sample.

Repeated Sampling and the Normal Distribution

In almost all cases, our estimates will be normally distributed (at least in large samples).

For example, we know that the sample mean of a variable is approximately normally distributed:

$$\bar{x} = \frac{1}{n}\sum_i x_i \sim \mathcal{N}(\mu, \frac{\sigma_x^2}{n})$$

Repeated Sampling and the Normal Distribution

In almost all cases, our estimates will be normally distributed (at least in large samples).

For example, we know that the sample mean of a variable is approximately normally distributed:

$$\bar{x} = \frac{1}{n}\sum_i x_i \sim \mathcal{N}(\mu, \frac{\sigma_x^2}{n})$$

This allowed us to do things like:

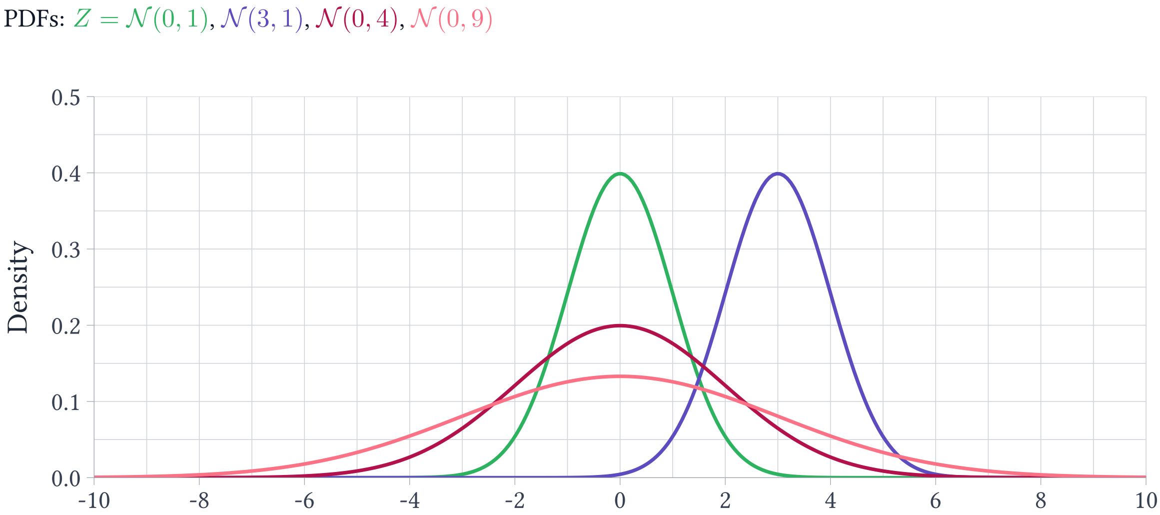
1. Form confidence intervals, and
2. Perform hypothesis tests

Normal Distribution

We say a random variable is normally distributed and write it as

$$x \sim \mathcal{N}(\mu, \sigma^2),$$

where $\mu = \mathbb{E}[x]$ and $\sigma^2 = \text{Var}\,(x)$


PDFs: $Z = \mathcal{N}(0,1)$, $\mathcal{N}(3,1)$, $\mathcal{N}(0,4)$, $\mathcal{N}(0,9)$
Density
0.0
0.1
0.2
0.3
0.4
0.5
-10
-8
-6
-4
-2
0
2
4
6
8
10

Properties of Normal Distribution

We have $x \sim \mathcal{N}(\mu, \sigma^2)$. Say we perform some transformation of x:

$$ax + b$$

The modified variable is still normally distributed. Using properties of expectation and variance, we have:

1. $\mathbb{E}[ax + b] = a\,\mathbb{E}[x] + b = a\mu + b$
2. $\text{Var}\,(ax + b) = \text{Var}\,(ax) = a^2\ \text{Var}\,(x) = a^2\sigma^2$

Properties of Normal Distribution

We have $x \sim \mathcal{N}(\mu, \sigma^2)$. Say we perform some transformation of x:

$$ax + b$$

The modified variable is still normally distributed. Using properties of expectation and variance, we have:

1. $\mathbb{E}[ax + b] = a\,\mathbb{E}[x] + b = a\mu + b$
2. $\text{Var}\,(ax + b) = \text{Var}\,(ax) = a^2\ \text{Var}\,(x) = a^2\sigma^2$

$$ax + b \sim \mathcal{N}\left(a\mu + b, a^2\sigma^2\right)$$

Random Vectors

We can extend this distribution to the joint distribution of multiple random variables. Let

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$ be a random vector.

These variables could be correlated with one another. For example, say x_1 is height and x_2 is weight of a surveyed person (or a sample mean of them).

- In repeated sampling, we expect x_1 and x_2 to be positively correlated.

Expectation of Random Vector

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

The expectation of the vector is just the expectation of each component variable

$$\mathbb{E}[x] = \begin{bmatrix} \mathbb{E}[x_1] \\ \vdots \\ \mathbb{E}[x_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$\tilde{x} = \begin{bmatrix} x_1 - \mu_1 \\ \vdots \\ x_n - \mu_n \end{bmatrix}$$

This vecotr has mean 0. We can write the covariance of the random vector as

$$\mathbb{E}\big[\tilde{x}\tilde{x}'\big] = \begin{bmatrix} \mathbb{E}[(x_1 - \mu_1)(x_1 - \mu_1)] & \dots & \mathbb{E}[(x_1 - \mu_1)(x_n \mu_n)] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[(x_n - \mu_n)(x_1 - \mu_1)] & \dots & \mathbb{E}[(x_n - \mu_n)(x_n \mu_n)] \end{bmatrix}$$

Covariance of Random Vector

From the previous slide, we have

$$\Sigma \equiv \mathbb{E}\big[(x-\mu)(x-\mu)'\big] = \begin{bmatrix} \mathsf{Var}\,(x_1) & \dots & \mathsf{Cov}\,(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathsf{Cov}\,(x_n, x_1) & \dots & \mathsf{Var}\,(x_n) \end{bmatrix}$$

The diagonal of this matrix is the variance of each variable. The off-diagonal elements are the covariance between the varaibles.

Covariance of Random Vector

From the previous slide, we have

$$\Sigma \equiv \mathbb{E}\big[(x-\mu)(x-\mu)'\big] = \begin{bmatrix} \mathsf{Var}\,(x_1) & \dots & \mathsf{Cov}\,(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathsf{Cov}\,(x_n, x_1) & \dots & \mathsf{Var}\,(x_n) \end{bmatrix}$$

The diagonal of this matrix is the variance of each variable. The off-diagonal elements are the covariance between the varaibles.

- If the off-diagonal element are all 0, then the random variables are uncorrelated.

Multivariate Normal Distribution

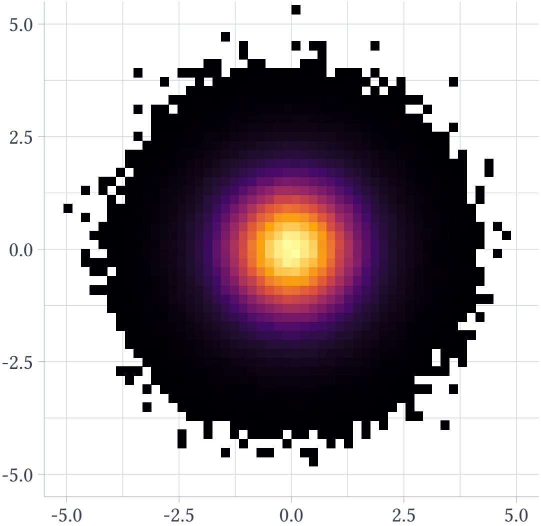
In particular, we might estimate a few parameters in our model and they will each be normally distributed.

But, even stronger, these estimates (a random vector) x are multivariate normally distributed:

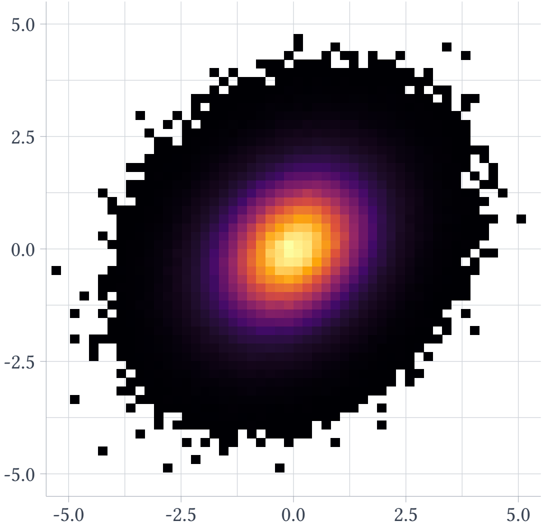
$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$$

where $\Sigma \equiv \mathbb{E}[(x-\mu)(x-\mu)']$

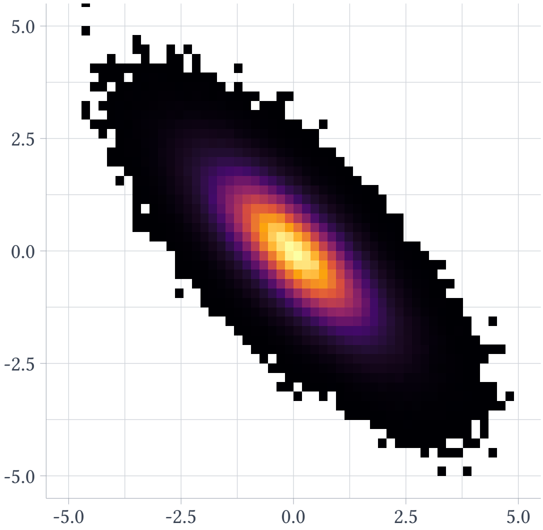
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \right)$$



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.75 \\ -0.75 & 1 \end{bmatrix}\right)$$



Linear combinations of multivariate random variable

Say $x \sim \mathcal{N}\left(\mu, \Sigma\right)$ and $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

Then, the product $a'x$ is a linear combination of x_i and is normally distributed:

$$a'x = \sum_i a_i x_i \sim \mathcal{N}(?, ?)$$

Linear combinations of multivariate random variable

Say $x \sim \mathcal{N}\left(\mu, \Sigma\right)$ and $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

First, consider the expectation of $a'x$

$$
\begin{aligned}
\mathbb{E}[a'x] = \mathbb{E}\left[\sum_i a_i x_i\right] &= \sum_i a_i \, \mathbb{E}[x_i] \\
&= \sum_i a_i \mu_i \\
&= a'\mu
\end{aligned}
$$

Linear combinations of multivariate random variable

Second, consider the variance of $a'x$

$$\begin{aligned}
\mathsf{Var}\left(a'x\right) &= \mathbb{E}\big[(a'x - a'\mu)(a'x - a'\mu)'\big] \\
&= \mathbb{E}\big[(a'(x-\mu))(a'(x-\mu))'\big]
\end{aligned}$$

Second, consider the variance of $a'x$

$$\begin{aligned}
\mathsf{Var}\left(a'x\right) &= \mathbb{E}\big[(a'x - a'\mu)(a'x - a'\mu)'\big] \\
&= \mathbb{E}\big[(a'(x-\mu))(a'(x-\mu))'\big] \\
&= \mathbb{E}\big[a'(x-\mu)(x-\mu)'a\big] = a'\Sigma a
\end{aligned}$$

Linear combinations of multivariate random variable

Second, consider the variance of $a'x$

$$\begin{aligned}
\text{Var}\left(a'x\right) &= \mathbb{E}\big[(a'x - a'\mu)(a'x - a'\mu)'\big] \\
&= \mathbb{E}\big[(a'(x - \mu))(a'(x - \mu))'\big] \\
&= \mathbb{E}\big[a'(x - \mu)(x - \mu)'a\big] = a'\Sigma a
\end{aligned}$$

Therefore, if $x \sim \mathcal{N}\left(\mu, \Sigma\right)$, then

$$a'x = \sum_i a_i x_i \sim \mathcal{N}(a'\mu, a'\Sigma a)$$

Matrix Fundamentals

Matrix Transformations

Inverse of a Matrix

Normal Random Variables

Derivatives of matrix expressions

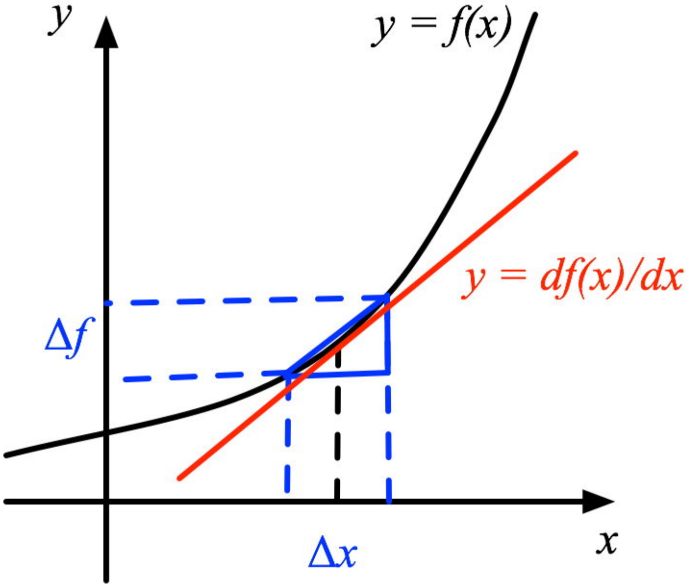
If you recall, we had the following definition of a derivative:

$$\frac{d}{dx}f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

You could write this as an approximation:

$$f(x + dx) - f(x) \approx \frac{d}{dx}f(x) * dx$$

In words, the change in f is approximately equal to the derivative times the change in x



y
y = f(x)
y = df(x)/dx
Δf
Δx
x

Taylor Expansion

This formulation is actually the basis of the Taylor Expansion that you may have learned in your calculus course:

$$f(x + dx) \approx f(x) + \frac{d}{dx} f(x) * dx + \text{something small}$$

- So long as dx is "small", this approximation is quite accurate.

Multivariate derivative

Say f is now a function that takes a vector x ($\mathbb{R}^n$) and produces a scalar output ($\mathbb{R}$). How f changes depends on which of the input x_i you are changing.

$$\frac{\partial}{\partial x} f(x) \equiv \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$$

This is called the "gradient" and it is a vector containing the partial derivatives

Multivariate derivative

Our linear approximation holds in this setting to. Let dx be the column vector of changes in each x_i: $dx = \begin{bmatrix} dx_1 & \dots & dx_n \end{bmatrix}'$.

Multivariate derivative

Our linear approximation holds in this setting to. Let dx be the column vector of changes in each x_i: $dx = \begin{bmatrix} dx_1 & \dots & dx_n \end{bmatrix}'$.

Then our approximation is as follows:

$$
\begin{aligned}
f(x + dx) &\approx f(x) + \frac{\partial}{\partial x} f(x)' dx \\
&= f(x) + \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) & \dots & \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}
\end{aligned}
$$

Multivariate derivative

Our linear approximation holds in this setting to. Let dx be the column vector of changes in each x_i: $dx = \begin{bmatrix} dx_1 & \dots & dx_n \end{bmatrix}'$.

Then our approximation is as follows:

$$
\begin{aligned}
f(x+dx) &\approx f(x) + \frac{\partial}{\partial x} f(x)' dx \\
&= f(x) + \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) & \dots & \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} \\
&= f(x) + \sum_i \frac{\partial f(x)}{\partial x_i} dx_i
\end{aligned}
$$

Useful matrix derivative rules

Let x and y be vectors of the same length. The derivative of the dot product $x'y$ can be given as

$$\frac{\partial}{\partial x} x'y = y$$

and

$$\frac{\partial}{\partial y} x'y = x$$

Useful matrix derivative rules

Let $\boldsymbol{A}$ be a matrix and x a vector.

$$\frac{\partial}{\partial x}\boldsymbol{A}x = \boldsymbol{A}$$

More, using the *chain rule* for derivatives, you could show

$$\frac{\partial}{\partial x}x'\boldsymbol{A}x = 2\boldsymbol{A}x$$

You can see the proofs here: `https://bookdown.org/compfinezbook/introcompfinr/Derivatives-of-Simple-Matrix-Functions.html`