1 Review of Key Concepts

1.1 Linear Algebra

Introduction

Linear algebra provides a way of representing and operating on sets of linear equations of the form:

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ 3x_1 + 4x_2 &= 6 \end{aligned} \tag{1}$$

The representation of these linear equations is done more compactly using matrices. A **matrix** is an array of numbers arranged in rows and columns and used to analyze linear equations, such as the one above. As an example, we can rewrite system of equations (1) more compactly as $AX = B$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \text{and } B = \begin{bmatrix} 5 \\ 6 \end{bmatrix}.$$

Besides the obvious space savings, there are many advantages to analyzing linear equations in this form when studying econometrics.

In the above example, A here is a matrix with two rows and two columns, and with entries 1, 2, 3 and 4. We say that A is a 2×2 or a 2 by 2 matrix. In general, we denote a matrix G with m rows and n columns by $G \in \mathbb{R}^{m \times n}$ where the entries of G are real numbers. We use the notation g_{ij} or G_{ij} or $G_{i,j}$ to denote the entry of G in the ith row and the jth column.

Matrix Multiplication

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times r}$ is the matrix $C = AB \in \mathbb{R}^{m \times r}$, where $C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$. In order for the matrix product to exist, the number of columns in A must equal the number of rows in B.

For a matrix-matrix multiplication, the (i, j)th entry of C is equal to the inner product of the ith row of A and the jth column of B. Specifically, we will have:

$$C_{i,j} = \begin{bmatrix} A_{i,1} & A_{i,2} & A_{i,3} & \dots & A_{i,n} \end{bmatrix} \begin{bmatrix} B_{1,j} \\ B_{2,j} \\ B_{3,j} \\ \vdots \\ B_{n,j} \end{bmatrix} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$$

Matrix multiplication follows the following properties:

- Matrix multiplication is associative: $(AB)C = A(BC)$
- Matrix multiplication is distributive: $A(B + C) = AB + AC$
- Matrix multiplication is not commutative in general, i.e. it can be the case that $AB \neq BA$

Matrix Operations and Properties

The **identity matrix**, denoted $I \in \mathbb{R}^{n \times n}$, is a square matrix with ones on the diagonal and zeros everywhere else, that is $I_{i,j} = 1$ if $i = j$ and $I_{i,j} = 0$ if $i \neq j$. The identity matrix has the property that for all $A \in \mathbb{R}^{n \times n}$, $AI = A = IA$.

The transpose of a matrix results from flipping the rows and the columns of the matrix. Given a matrix $A \in \mathbb{R}^{m\times n}$, its transpose $A^T \in \mathbb{R}^{n\times m}$ or $A^{'} \in \mathbb{R}^{n\times m}$ is the $n \times m$ matrix whose entries are given by:

$$(A^T)_{ij} = A_{ji}$$

Transposes follow the following properties:

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A + B)^T = A^T + B^T$

Rank and Linear Dependency

A set of vectors $\{x_1, x_2, \ldots, x_n\}$ is said to be linearly independent if no vector can be represented as a linear combination of the remaining vectors. If one vector belonging to this set can be represented by a linear combination of the remaining vectors, then the vectors are said to be linearly dependent. If $x_n = \sum_{i=1}^{n-1} k_i x_i$ for some scalar values $k_1, k_2, \ldots, k_{n-1} \in \mathbb{R}$, then the vectors $\{x_1, x_2, \ldots, x_n\}$ are linearly dependent. Otherwise, they are linearly independent.

The **column rank** of a matrix $A \in \mathbb{R}^{m\times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set. In the same way, the **row rank** is the largest number of rows of A that constitute a linearly independent set. For any matrix $A \in \mathbb{R}^{m\times n}$, the column rank of A is equal to the row rank of A, and both quantities are referred to collectively as the **rank** of A, denoted as rank(A).

The rank follows the following properties:

- For $A \in \mathbb{R}^{m\times n}$, rank($A$) $\leq$ min(m,n). If rank(A) = min(m,n), then A is said to be **full rank**.
- For $A \in \mathbb{R}^{m\times n}$, rank($A$) = rank($A^T$)
- For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times r}$, rank($AB$) $\leq$ min(rank(A), rank(B))
- For $A, B \in \mathbb{R}^{m\times n}$, rank($A + B$) $\leq$ rank(A) + rank(A)

Inverse Matrix

The **inverse** of a square matrix $A \in \mathbb{R}^{n\times n}$ is denoted A^{-1} and is the unique matrix such that $A^{-1}A = I = AA^{-1}$, where I is the identity matrix.

Not all matrices have inverses. Non-square matrices do not have inverses by definition. A square matrix that does not have a matrix inverse is referred to as a **singular matrix**. In order for a square matrix A to have an inverse A^{-1}, A must be full rank. The following are properties of the inverse matrix. For invertible matrices $A, B \in \mathbb{R}^{n\times n}$:

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^T = (A^T)^{-1}$

For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we can find the inverse directly using the formula:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Note that A^{-1} exists only when $ad - bc \neq 0$.

1.2 Probability and Statistics

Random Variables and Distributions

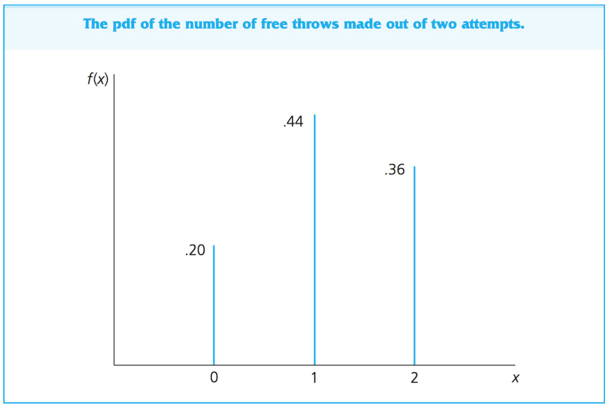
In essence, a **random variable** is a function that maps a sample space of possible outcomes to probabilities. It can be **discrete** where there are a countable number of distinct possible values (e.g. number of completed years of school) or **continuous** where there are uncountable possible values (e.g. a person's height). Once a random variable is drawn from the distribution, it becomes the **realization** of a random variable.

Any discrete random variable can be completely described by detailing the possible values it takes, as well as the associated probability that it takes each value. The **probability density function** (**PDF**) of a discrete random variable X summarizes the information concerning the possible outcomes of X and the associated probabilities. [1]

$$f(x_j) = P(X = x_j), \;\; j = \{1, 2, 3, 4, 5, ...k\}$$

Below, we show the PDF of the number of throws made out of two attempts by professional basketball players:

$$f(0) = 0.20 \,; f(1) = 0.44 \,; f(2) = 0.36$$

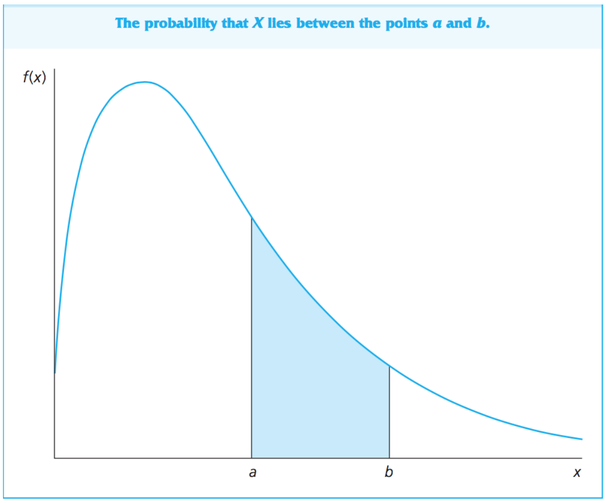


We can define a probability density function for continuous variables as well. However, it doesn't make sense to talk about the probability that a continuous random variable takes on a particular value,[2] rather the PDF computes the probability of events involving a certain range. The probability that X takes on a value within the interval $[a, b]$ is given by

$$Pr(a < X < b) = \int_a^b f(x)dx$$

[1]This is sometimes called probability mass function for discrete random variables.

[2]The idea is that a continuous random variable X can take on so many possible values that we cannot count them or match them up with the positive integers, so logical consistency dictates that X can take on each value with probability zero. (Woolridge p.717)



When computing probabilities for continuous random variables, it is easiest to work with the **cumulative distribution function (CDF)**. The CDF of a random variable is defined as:

$$F(x) = Pr(X \leq x)$$

For discrete random variables, this is obtained by summing the PDF over all values x_j such that $x_j \leq x$. For a continuous random variable, $F(x)$ is the area under the PDF, $f(x)$, to the left of the point x. Two important properties of CDF's that we will use later in the course:

$$Pr(X > c) = 1 - F(c)$$
$$Pr(a < X \leq b) = F(b) - F(a)$$

Joint Distributions, Conditional Distributions, and Independence

For a pair of random variables X, Y, the **joint cumulative distribution function (joint CDF)** F_{XY} is given by

$$F_{X,Y} = P(X \leq x, Y \leq y).$$

The **joint probability density function (joint PDF)** of two discrete random variables X, Y is

$$p_{X,Y}(x, y) = Pr(X = x, Y = y).$$

The joint PDF of two continuous random variables X, Y is

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}.$$

We might also be interested in establishing how X varies with different values of Y: this is the conditional distribution of Y given X, which is described by the **conditional probability density function (conditional PDF)**. For discrete random variables, the conditional PDF of Y given $X = x$ can be written according to its definition as:

$$p_{Y|X}(y|x) \equiv Pr(X = x|Y = y) = \frac{Pr(X = x, Y = y)}{Pr(X = x)}.$$

Similarly for continuous random variables, the conditional PDF of X given $Y = y$ can be written as:

$$f_{(X|Y)}(x|y) = \frac{f_{Y,X}(y, x)}{f_Y(y)}$$

where $f_{Y,X}(y, x)$ is the joint PDF of Y and X, while $f_Y(y)$ is the marginal density of Y.

Two variables are **independent** if the joint PDF is equal to the product of the individual variables' PDF.

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y), \forall x, y.$$
$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \forall x, y.$$

where $f_X(x)$ and $f_Y(y)$ are the marginal density functions for X and Y.

Features of Probability Distributions

Expected Value

If X is a random variable, the **expected value** (or expectation) of X, denoted E(X), is the weighted average of all possible values of X. The weights are determined by the probability density function. The expected value is also called the population mean. Formally, if X is discrete with a finite list of possible outcomes,

$$\mathbb{E}(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_k f(x_k) = \sum_{j=1}^{k} x_j f(x_j)$$

If X is continuous and has a PDF $f(\cdot)$ on the real number line,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) d(x)$$

Now for a quick example:

x_j	$p(X = x_j)$
-1	1/8
0	1/2
2	3/8

$$\mathbb{E}(X) = (-1)(1/8) + 0(1/2) + 2(3/8) = 5/8$$

Variance and Standard Deviation

The **variance** tells us the expected distance from X to its mean:

$$\begin{aligned} Var(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] \\ &= \sum_{j=1}^{k} f(x_j)(x_j - \mathbb{E}(X))^2 \text{ for discrete case,} \\ &= \int_{-\infty}^{+\infty} f(x)(x - \mathbb{E}(X))^2 dx \text{ for continuous case.} \end{aligned}$$

Going back to our example:

$$Var(X) = \frac{1}{8}(-1 - \frac{5}{8})^2 + \frac{1}{2}(0 - \frac{5}{8})^2 + \frac{3}{8}(2 - \frac{5}{8})^2$$

Note the squaring eliminates the sign from the distance measure; the resulting positive value corresponds to the notion of distance, and treats values above and below symmetrically.

The **standard deviation** of a random variable, denoted sd(X) is the positive square root of the variance: $sd(X) = +\sqrt{Var(X)}$

Note

$$\begin{aligned} Var(aX + b) &= a^2 Var(X) \\ sd(aX + b) &= a \cdot sd(X) \end{aligned}$$

This last property makes the standard deviation more natural to work with than the variance. As an example, take a random variable X measured in dollars. Next define Y=1000X. Suppose $\mathbb{E}(X) = 20$ and $sd(X) = 6$. Then:

$$\begin{aligned} \mathbb{E}(Y) &= 1000\mathbb{E}(X) = 20,000 \\ sd(Y) &= 1000 sd(X) = 6,000 \\ Var(Y) &= (1000)^2 sd(X) = 6,000,000 \end{aligned}$$

The expected value and the standard deviation both increase by the same factor 1,000, whereas the variance of Y scales by 1,000,000.

Covariance and Correlation

If X and Y are two random variables, with expectations μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively, then their **covariance** is as follows:

$$Cov(X,Y) = \sigma_{XY} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)],$$

and their **correlation** is as follows:

$$Corr(X,Y) = \rho_{XY} = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{sd(X)sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Some useful properties of covariance are as follows, for random variables X, Y, and Z, and two constants a and b,

$$\begin{aligned}
Cov(X,Y) &= Cov(Y,X), \\
Cov(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y], \\
Var(X) &= Cov(X,X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \\
Cov(aX,Y) &= aCov(X,Y), \\
Cov(X+Z,Y) &= Cov(X,Y) + Cov(Z,Y).
\end{aligned}$$

2 Review of Select Quiz Questions

2.1 Linear Algebra

1. Consider the matrix
$$A = \begin{bmatrix} 1 & b \\ 0 & c \end{bmatrix}$$
What is the transpose of A (denote the transpose of a matrix A by A')?

2. Obtain explicitly the matrix $C = A'A$

3. Suppose for this problem that $b = 0$ and $c = 1$. What is the rank of the matrix A? Find the inverse of this matrix.

4. Consider now the matrix A but now with $c = 0$ and $b = 0$. What is the rank of this matrix? Is it invertible? What is the relationship between the rank of a square matrix and the existence of its inverse?

5. Consider the following equations and expressions

$$\beta_2 = 0$$

$$\beta_1 + \beta_3 = 3$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

   Rewrite these equations in matrix form (i.e. of the form $A\beta = b$ for some matrix A that you need to find and the matrix β defined above and for some 2×1 vector b that you need to find as well).

6. Consider the two $m \times 1$ vectors given by

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

   What is $\mathbf{a}'\mathbf{b}$? What is $\mathbf{b}'\mathbf{a}$?

7. Consider the $r \times m$ matrix $\mathbf{A}$ written in terms of its m columns

$$\mathbf{A}_{r \times m} = \begin{pmatrix} \mathbf{a}_1 & . & . & \mathbf{a}_m \end{pmatrix}$$

where $\mathbf{a}_j$ is a $r \times 1$ vector and denotes the j^{th} column of $\mathbf{A}$. Consider the $m \times d$ matrix $\mathbf{B}$ expressed in terms of its rows

$$\mathbf{B}_{m\times d}=\begin{pmatrix} \mathbf{b}_1 \\ \cdot \\ \cdot \\ \mathbf{b}_m \end{pmatrix}$$

where $\mathbf{b}_j$ is a $1 \times d$ vector that denotes the j^{th} row of $\mathbf{B}$.

(a) What is the dimension of the matrix $\mathbf{AB}$

(b) Obtain an expression for $\mathbf{AB}$ in terms of the vectors above.

2.2 Probability and Statistics

1. Consider a random variable X that takes on the value 0 with probability p and the value 1 with probability $1 - p$

   (a) Compute the expectation of X, denoted by $\mathbb{E}(X)$

   (b) Compute the variance of X, denoted by $Var(X)$

   (c) Suppose you define a new random variable

   $$Z = 1 + 2X$$

   Repeat parts (a) and (b) for the random variable Z.

   (d) Compute the covariance between X and Z.

2. Let the random variable X denote the outcome of the toss of a six sided die. That is, X can take on the values $\{1, 2, 3, 4, 5, 6\}$. Suppose that the die is fair so that each number in $\{1, 2, 3, 4, 5, 6\}$ has an equal probability of occurring.

(a) Consider the events $A = \{1,3,5\}$, $B = \{1,4\}$, $C = \{2,4,6\}$ and $D = \{3,5\}$ and compute their probabilities.

(b) Compute the probabilities $\mathbb{P}(A \cap B)$ and $\mathbb{P}(A \cup B)$.

(c) Are the events A and C independent?

(d) Compute the conditional probability $\mathbb{P}(D|A)$

3. Consider two random variables X and Y each of which can take on two values $\{0,1\}$. The joint distribution is given by

$$\mathbb{P}(X = 0, Y = 0) = p$$
$$\mathbb{P}(X = 0, Y = 1) = q$$
$$\mathbb{P}(X = 1, Y = 0) = r$$

(a) Express $\mathbb{P}(X = 1, Y = 1)$ as a function of (p, q, r)

(b) Express the covariance between X and Y as a function of (p, q, r). Is the covariance between X and Y the same as the covariance between Y and X?

(c) Express the correlation between X and Y as a function of (p, q, r)

(d) Compute the expectation of Y, $\mathbb{E}(Y)$ this is also sometimes called the unconditional expectation of Y.

(e) Compute the conditional expectation of Y given that X is equal to 1 $\mathbb{E}(Y|X = 1)$. That is, what would you expect the average value of Y to be given that X is equal to 1.