MAT217Lectures

Mark A. Carty
Scottsdale Community College

November 19, 2008

§3.1 Systems of Equations

A system of two linear equations in two variables is any problem expressible in the form

$$\begin{cases} ax + by = h \\ cx + dy = k \end{cases}$$

where x and y are the variables and the constants a, b, c, d, h, k are such that at least one of coefficients a, b, c, d is not zero.

§3.1 Solving linear system of equations

A solution of a system of linear equations in two variables is a pair of values for the variables that satisfy all the equations. We can solve systems of linear equations by three different methods: graphing, substitution, and elimination.

§3.1 Graphical Method

The graph of a pair of equations may take three different forms: two lines intersecting at just one point (a unique solution), two lines that don't intersect (no solutions), or two lines that are the same (infinitely many solutions). For lines that don't intersect, we say that the equations are inconsistent, and for the lines that are the same we call the equations dependent.

§3.1 Using Graphical Method

Sketch the graphs of $2x + 3y = 12$ and $2x - 3y = 0$. To graph $2x + 3y = 12$ we find the intercepts by setting each variables in turn to equal to zero. We should get $(0, 4)$ for the y-intercept and $(6, 0)$ for the x-intercept. Then draw a line through both points. We repeat the same process for the other equation.

3.1 Subsitution Method

The graphical method is limited b the accuracy of your sketch. Another way of solving a system of linear equations is called subsitution method and involves solving one of equations for a variable and then substituting the result into the other equation.

§3.1 Elimination Method

The elimination method attempts to remove one variable from each and write them as an equivalent form

$$\begin{cases} x + 0y = p \\ 0x + y = q \end{cases}$$

so that the solutions can read off as $x = p, y = q$. If this cannot be done, the method will identity the system as inconsistent or dependent.

§3.2 Matrices

A matrix is a rectangular array of elements. The rectangular array has row (with first at the top, the second below the first, and so on) and columns (with the first on the left, second to right of the first, and so on).

The dimension of a matrix is a product of the number of rows and columns. A 2 by 2 matrix has dimension 2×2.

§3.2 Matrices

$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ Row matrix of dimension 1×3

$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ Column matrix of dimension 3×1

$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ Matrix of dimension 3×3

§3.2 Augmented Matrix

The system of linear equations below

$$\begin{cases} ax + by &= h \\ cx + dy &= k \end{cases}$$

can be represented by a matrix that is called an augmented matrix.

$$\left(\begin{array}{cc|c} a & b & h \\ c & d & k \end{array} \right).$$

Notice that the first column has entries that are the x coefficients of two equations and the second column has entries that are the y coefficients of two equations. The constant terms are to the right of the vertical line.

§3.2 Elementary row operations

We can solve a system of linear equations by operating on the augmented matrix, using elementary row operations to obtain reduced row echelon matrix. This matrix is identity matrix with a column matrix augmented to it. That is,

$$\left(\begin{array}{cc|c} 1 & 0 & p \\ 0 & 1 & q \end{array}\right).$$

Elementary Matrix Operations:

- Switch any two rows.
- Multiply or divided one of the rows by a nonzero number.
- Replace row by its sum or different with a multiple of another row.

§3.3 The Gauss-Jordan Method

The Gauss-Jordan Method is a variation of Gaussian elimination. The Gauss-Jordan method is used to solve large systems of equations numerically.

§3.3 Gauss-Jordan Method Con't

Gauss-Jordan Method:

- If any rows have leading zeros, switch the rows so that the one with the fewest leading zeros are the the top, down to the ones with the most leading zeros at the bottom.
- In the first row, find the leftmost nonzero entry and divide the row through by that number. This gives a leftmost 1 in that row.
- Add or subtract multiples of the first row to each other row to obtain zeros in the rest of the column above and below the leftmost 1 found in step.
- Repeat steps 1, 2, and 3 but replace "first row" by "second row" and then by "third row" and so on. Stop when you reach the bottom row or a row consisting entirely of zeros–when this happens the matrix is row-reduced.

§3.4 Ring

A ring is a set R equipped with two binary operations $+ : R \times R \to R$ and $\cdot : R \times R \to R$:

- $(R, +)$ is an abelian group with identity element 0, meaning that for all $a, b, c \in R$, the following axioms hold:
  - $(a + b) + c = a + (b + c)$ (+ is associative)
  - $0 + a = a$ (0 is the identity)
  - $a + b = b + a$ (+ is commutative)
  - for each $a \in R$ there exists $-a \in R$ such that $a + (-a) = (-a) + a = 0$ ($-a$ is the inverse element of a)
- $(R, \cdot)$ is a monoid with identity element 1, meaning that for all $a, b, c \in R$, the following axioms hold:
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ ($\cdot$ is associative)
  - $1 \cdot a = a \cdot 1 = a$ (1 is the identity)
- Multiplication distributes over addition:
  - $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
  - $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

§3.4 Matrix algebra

The set of all m by n matrices with entries in $\mathbb{R}$ is a ring. However, multiplication is not commutative. That is,

$$A \cdot B \neq B \cdot A.$$

The product of a scalar (a number) times a matrix, kA, is the matrix A with each element multiplied by k. For a two matrices A

and B of the same dimension, their sum $A + B$ and difference $A - B$ are found by adding or subtracting corresponding elements. Matrices of different dimensions cannot be added or subtracted.

§3.4 Matrix Multiplication

If A is an $m \times p$ matrix and B is an $p \times n$ matrix, the matrix product $A \cdot B$ is the matrix $m \times n$ whose elements in row i and column j is the product of the row i from A times column j from B. If the row-length of A does not equal the column length of B then the product $A \cdot B$ is no defined.

§4.1 Linear Inequalities

A typical linear programming problem asks for the maximum or minimum value of a linear function subject to a collection of linear inequalities. Linear programming problems often arise in modern business, sometimes involving many variables, and are usually solved by a procedure known as the simplex method. The simplex method is a numerical algorithm based on the geometry of linear.

§4.1 Graphing linear inequalities

- Draw the boundary line (possibly using the intercepts)
- Use a test point not on the line (possibly the origin) to determine the correct side. If this point satisfies the inequality, the point lies on the correct side; if not, the other side is correct side.
- Shade the correct side to show the feasible region.

§4.2 How to Solve a Linear Programming Problem

If the region is bounded, list the vertices and calculate the value of the objective function at each. The solution occurs at the vertex that gives the largest (or smallest) value.

If the region is unbounded, first check whether the objective function improves in an unbounded direction. If it does, there is no solution; if it does not, list the vertices and calculate the value of the objective function at each; the solution occurs at the vertex that gives the largest (or smallest) value.

§4.2 Determine Whether a Solution Exists

If the feasible region is unbounded, the linear programming problem say may not have a solution. First, determine the directions in which the feasible region is unbounded.

- If the objective function improves when the x and y variables take values further out an unbounded direction, then the solution does not exist.
- If the objective function does not improve when the x and y variables take values further out in any of the unbounded directions, then the solution does exist.

§5.1 Probability

A set is any well-defined collection of objects. To say that a set is well defined, we mean you can tell whether an object is in the set or not.

The intersection of sets A and B, represented by $A \cap B$, is the set of all elements that are in both A and B.

The union of sets A and B, represented by $A \cup B$, is the set of all elements that are in either A or B (or both).

The complement of a set A, represented by A^c, is the set of all elements not in A.

Two sets A and B are said to be disjoint when both sets have no elements in common.

§5.3 Axioms of Probability Theory

- Let A be an event. Then $0 \leq P(A) \leq 1$.
- Let S be the sample space. Then $P(S) = 1$.
- $P(\emptyset) = 0$.
- If A and B are two mutually exclusive events then, $P(A \cup B) = P(A) + P(B)$.
- If $A \subset B$, then $P(B) = P(B \setminus A) + P(A)$.

§5.3 Some Propositions Probability Theory

Let A and B sets. Then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. It follows that

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

$P(A^c) = 1 - P(A)$.

Proof. $1 = P(S) = P(A \cap A^c) = P(A) + P(A^c)$.

§5.4 Conditional Probability and Independence

For events A and B with $P(B) > 0$, the conditional probability of A given B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

§5.4 Conditional Probability and Independence

Events A and B are independent if

$$P(A \cap B) = P(A) \cdot P(B)$$

Events that are not independent are dependent.

§5.5 Bayes' Formula

$$P(B_k \mid A) = \frac{P(A \mid B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)}$$

where A is a set and $B_1, ..., B_n$ are pairwise disjoint sets.

Example: In a town, 60% of citizens are Republicans and 40% are Democrats. In the last elections 55% of Republicans voted and 65% of Democrats voted. If a voter is randomly selected, what is the probability that the person is a Republican?

§5.5 Solution to Example

Solution: Let B denote the party affiliation and B_1 and B_2 are Republicans and Democrats, respectively. Let A denote those that voted in the last election. We want to find $P(B_1 \mid A)$. By Bayes Rule, it follows that

$$\begin{aligned} P(B_1 \mid A) &= \frac{P(A \mid B_1) \cdot P(B_1)}{P(A \mid B_1) \cdot P(B_1) + P(A \mid B_2) \cdot P(B_2)} \\ &= \frac{(0.55)(0.60)}{(0.55)(0.60) + (0.65)(0.40)} \\ &= 0.56. \end{aligned}$$

§5.6 Random Variables and Distributions

Def: A random variable X is an assignment of a number to each element in the sample space. The probability distribution of the random variable X is the collection of all probabilities $P(X = x)$ for each possible value x.

Example: A coin is tossed six times. Find and graph the probability distribution for number of tails.

Solution: The probability distribution is

$$\begin{aligned} P(X = x) &= \binom{6}{x}(0.5)^x(0.5)^{6-x} \\ &= \binom{6}{x}(0.5)^6. \end{aligned}$$

§5.6 Expected Value

A random variable X taking values $x_1, x_2, \ldots, x_n$ with probability $p_1, p_2, \ldots, p_n$ has expected value:

$$E(X) = \sum_{k=1}^{n} x_k p_k.$$

§Supplementary Notes on Differentiation

Definition: Let $U \subseteq \mathbb{R}$ be open, $f : U \to \mathbb{R}$, and $x_0 \in U$. f is differentiable at x_0 if

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. In this case we denote the value of the limit by $f'(x_0)$, and call it the derivative of f at x_0.

§Some Important concepts

Suppose $f, g : U \to \mathbb{R}$ are differentiable at $x_0 \in U$. Then

(i) $(f+g)'(x_0)$ exists and equals $f'(x_0) + g'(x_0)$

(ii) $(fg)'(x_0)$ exists and equals $f'(x_0)g(x_0) + f(x_0)g'(x_0)$

(iii) If $g(x_0) \neq 0$, then $(f/g)'(x_0)$ exists and equals

$$\frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$$

§Chain Rule

Suppose $f : U \longrightarrow \mathbb{R}$ is differentiable at $x_0 \in U$ and $g : V \longrightarrow \mathbb{R}$ is differentiable at $f(x_0) \in V$. Then $g \circ f$ is differentiable at x_0, and

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

§Continuous Random Variables

We say that X is a continuous random variable if there exists a nonnegative function f, defined for all real numbers having the property that for any set A of real numbers

$$P(X \in A) = \int_A f(x)dx.$$

The function is called the probability density function of the random variable X.

§Cumulative Distribution Function

The relationship between the cumulative distribution F and the probability density f is expressed by

$$F(x) = P\{X \in (-\infty, x]\} = \int_{-\infty}^{x} f(t)dt$$

Differentiating both sides of the above yields

$$\frac{d}{dx}F(x) = f(x).$$

§Uniform Random Variable

We say that X is a uniform random variable on the interval (a, b) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

The cumulative distribution is given by

$$F(x) = \begin{cases} 0 & \text{if } x \in (-\infty, a] \\ \frac{x-a}{b-a} & \text{if } x \in (a, b) \\ 1 & \text{if } x \in [b, \infty) \end{cases}$$

§Normal Random Variable

We say that X is normal random variable, or simply that X is normally distributed, with parameters μ and σ^2 if the density of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \text{ whenever } x \in \mathbb{R}$$

It follows that the distribution function of X can be expressed as

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

§Expectation of Continuous and Random Variables

We defined the expected value of a discrete random variable X by

$$E[X] = \sum_x xP\{X = x\}$$

If X is a continuous random variable having probability density function $f(x)$, then as

$$f(x)dx \approx P\{x \leq X \leq x + dx\} \text{ for } dx \text{ small}$$

it is easy to see that the analogous definition is to define the expected value of X by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

§Exponential Random Variables

A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

is said to be an exponential random variable (or, more simply, is said to be exponentially distributed) with parameter λ.

§Cdf of Exponential Random Variable

The cumulative distribution function (or cdf) $F(a)$ of an exponential random variable is given by

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= 1 - e^{-\lambda x}, \text{ where } x \geq 0. \end{aligned}$$

Note that

$$F(\infty) = \int_0^\infty \lambda e^{-\lambda x} dx = 1,$$

as, of course, it must, The parameter λ will now be shown to equal the reciprocal of the expected value.

§Memoryless Random Variable

We say that a nonnegative random variable X is memoryless if

$$P\{X > s + t \mid X > t\} = P\{X > s\} \text{ for all } s, t \geq 0$$

The condition is equivalent to

$$P\{X > s + t\} = P\{X > s\}P\{X > t\}$$

§Exponential Random Variables are Memoryless

Let X be an exponentially distributed random variable with the following density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Let $s, t \geq 0$. Then

$$\begin{aligned} P\{X > s + t\} &= 1 - F(s + t) \\ &= e^{-\lambda(s+t)} \\ &= e^{-\lambda s} e^{-\lambda t} \\ &= (1 - F(s))(1 - F(t)) \\ &= P\{X > s\} P\{X > t\}. \end{aligned}$$

And so, exponentially distributed random variables are memoryless.

§6.1 Statistics

A random sample is a selection of members of the population satisfying two requirements:

1. Every member of the population is equally likely to be included in the sample.
2. Every possible sample of the same size from the population is equally likely to be chosen.

§6.1 Statistics

There are four levels of data measurement:

Data	Uses
Nominal	Classify by name
Ordinal	Compare order (1st, 2nd,...)
Interval	Compare order and difference
Ratio	Compare order, differences, and ratios

§6.2 Statistics

Definition: The *mode* of a collection of data is the most frequently occurring value.

Definition: The *median* is the middle value of a list of data values when sorted in ascending or descending order.

Definition: The *mean* $\overline{x}$ of the n values $x_1, x_2, x_3, \ldots, x_n$ is

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

§6.3 Statistics

There are three ways to measure the spread of a data set. The simplest is the range, which is the largest value minus the smallest value. The range gives no indication of the typical variation away from the mean, just the difference between the extremes.

A box-and-whisker plot graphically shows the range of each quarter of the data and can be drawn easily using a graphing calculator.

The sample standard deviation measures the typical spread of the data the mean $\overline{x}$ as a single number and can be found easily using a calculator:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

§6.4 Normal Distribution

The normal probability distribution with mean μ and standard deviation σ is given by the curve

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad z \in \mathbb{R}$$

A random variable with this distribution is called a normal random variable with mean μ and standard deviation σ.

§6.4 z-Score

The z-score gives the number of standard deviations the value is from the the mean. It is defined as

$$z = \frac{x - \mu}{\sigma}.$$

§6.4 Examples

Let X be a normal random variable with mean $\mu = 12$ and standard deviation $\sigma = 3$. Find each probability as an area under the normal curve.

1. $P(9 \leq X \leq 15)$
2. $P(10.5 \leq X \leq 14)$

Solution: First, find the z-scores for the left and right endpoints in both problems. That is,

$$\begin{aligned} z_l &= \frac{9-12}{3} = -1 \\ z_r &= \frac{15-12}{3} = 1 \end{aligned}$$

Note: $\Phi(-a) = \Phi(a)$ because of symmetry. But then,

$$P(9 \leq X \leq 15) = 2\Phi(1) = 0.6426$$

§6.4 Examples

For problem 2, we see that

$$z_l = \frac{10.5 - 12}{3} = -0.5$$
$$z_r = \frac{14 - 12}{3} = 0.667$$

It follows that $\Phi(-0.5) = \Phi(0.5) = 0.1915$ and $\Phi(0.667) = .2422$. But then,

$$P(10.5 \leq X \leq 14) = \Phi(0.667) + \Phi(0.5) = 0.4337$$