Exercise 1 - Mathematical Introduction

Guy Singer

January 3, 2024

Part I
Basic Calculus

1 Derivatives, Summation and Integrals

The derivative is the slope of a function in any given point, and defined as $f'(x) = \lim_{h\to 0} \frac{f(x-h)-f(x)}{h}$. for example, the derivative of $f(x) = x^3$ with respect to x, written as $\frac{df}{dx}$, is $3x^2$. The derivative of $f(x) = x^3 + 2$ with respect to x is also $3x^2$. Here we see that constants, which are not dependent on x, don't affect the result of our derivation. The reason is clear if we think of the derivative as the slope of a line - the slope isn't affected by this constant, since it serves as an "offset" to the entire line, while not affecting the "rate" in which the function changes, which is exactly its slope. The general formula to derive a polynomial function is:

$$f(x) = x^n + C \Rightarrow f'(x) \equiv \frac{df}{dx} = nx^{m-1} \tag{1}$$

While we don't have to know how to use the derivative operator on every function $f(x)$ for this course, we're reminding ourselves of it here since it's very important for our understanding of integrals.

An indefinite integral can also be called an *antiderivative*, since the operation it does is to find the function that its derivation was the function we have in front of us:

$$f(x) = \int f'(x)dx \tag{2}$$

Above we define $f'(x)$ to be the derivative of $f(x)$. Notice that we're actually not being precise here. As we've seen in our first example, any constant added to $f(x)$ will result with the same derivative $f'(x)$. Thus, we should actually write:

$$\int f'(x)dx = f(x) + C \tag{3}$$

where C is a constant. The dx signs marks that this antidervative operation is done with respect to x. What makes this integral *indefinite* is the fact that it has no limits.

By reverse engineering the polynomial derivative formula we can deduce the polynomial integral formula:

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C \tag{4}$$

One more important property of the derivative and integral operations is *linearity* . A general operator $L(x)$ if:

$$L(ax) = aL(x) and L(x+y) = L(x) + L(y) \tag{5}$$

This property will be discussed more elaborately later in the course.

1.1 Question

What is the indefinite integral of $f(x) = 2x^{-3}$?

1.2 Answer

Remember we can take multiplicative constants outside of the integral:

$$\int 2x^{-3} dx = 2\int x^{-3} dx = 2\left(\frac{x^{-2}}{-2}\right) + C = -x^{-2} + C \tag{6}$$

Check yourself - derive the final expression and see how we receive the original expression we started with.

This "rule" has an exception for a certain power of x. Can you guess what it is? When $f'(x) = x^{-1}$, the integral is actually $f(x) = \ln|x| + C$.

For more derivative and integral formulas, look at the table in moodle.

1.3 Summation and definite integrals

A definite integral is the summation of the area under the curve of the original function between the specified limits (figure 1). This is also called the second fundamental theorem of calculus (first one being that an indefinite integral is an antiderivative). If you think about it, it's very surprising, but we won't go into the proof here.

The integral can be defined as the limit of summation of the area under the curve, when the width of every summed rectangle is infinitesimally small (figure 2).

We'll see how we denote this idea using summation: Let's find the integral between a and b of some function $f(x)$. To do that we'll divide the interval $[a, b]$ to n sub-intervals with length $\delta_i \equiv x_i - x_{i-1}$, where i goes from 1 to n. We'll

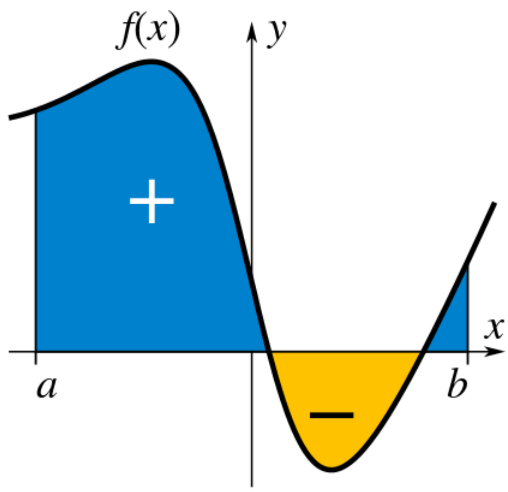

Figure 1: Definite integral: $\int_a^b f(x)dx$. The + and - signs indicate the way to perform the summation of the area under the curve. From Wikipedia.

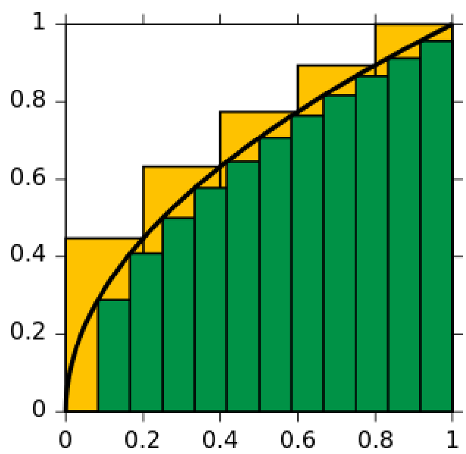

Figure 2: Summation of small sections along the x-axis. From Wikipedia.

also call t_i a point on the x-axis in the interval $[x_{i-1}, x_i]$. The definition of the integral is (informally) given by:

$$\int_a^b f(x)dx \equiv \lim_{\delta_i \to 0} \sum_{i=1}^{n} f(t_i)\delta_i = F(b) - F(a) \tag{7}$$

where $F(x)$ is the antiderivative of $f(x)$. Make sure to go over every term here and see that you fully understand it.

1.3.1 Question

What is the definite integral of $f(x) = 2\sin(x + pi)$ between 0 and $3\pi/2$?

1.3.2 Answer

We'll follow the definitions we have:

$$\begin{aligned} \int_0^{\frac{3\pi}{2}} 2\sin(x+\pi)\,dx &= 2\int_0^{\frac{3\pi}{2}} \sin(x+\pi)\,dx \\ &= 2\left[-\cos(x+\pi)\right]_0^{\frac{3\pi}{2}} \\ &= -2[0-(-1)] = -2 \end{aligned} \tag{8}$$

2 Complex Numbers

Sets of numbers:

- $\mathbb{N}$ - Natural numbers - 0, 1, 2, 3,...
- $\mathbb{Z}$ - Integers - ..., -2, -1, 0, 1, 2, ...
- $\mathbb{Q}$ - Rational - Ratios of two integers.
- $\mathbb{R}$ - Real - All numbers on the continuous number line, with no gaps. Includes, for example, $\sqrt{2}$ and π.
- $\mathbb{C}$ - Complex numbers - An imaginary number with (or without) a real part.

We define the imaginary unit as $i^2 \equiv -1$. i appears in many places in Math, Physics and Engineering, and it's more useful to our day-to-day lives than we think.

2.1 Question

What is i^{302}?

2.2 Answer

$$
\begin{aligned}
i^{302} = i^{300} \cdot i^2 = (i^3)^{100} \cdot (-1) &= -(i^2 \cdot i)^{4^{25}} \\
&= -\left[(-i)(-i)(-i)(-i)\right]^{25} \\
&= -[1]^{25} = -1
\end{aligned}
\tag{9}
$$

note that $-i$ is also a square root of -1, but i is considered the *principal root*, just like 3 is the principal root of 9, with -3 also being a square root of 9.

Using simple arithmetics we can also simplify square roots of negative numbers:

2.3 Question

What is $\sqrt{-52}$?

2.4 Answer

$$\sqrt{-52} = \sqrt{-1 \cdot 52} = i\sqrt{4 \cdot 13} = 2i\sqrt{13} \tag{10}$$

2.5 The Comlex Number

Complex numbers are additions of an imaginary number and a real number: $z = a + bi$. The real part of z, denoted as $\mathrm{Re}\{z\}$, is a, with its imaginary part $\mathrm{Im}\{z\}$ being b. A very useful way of looking at them is using the x- and y-axis for visualization in the complex plane (figure 3). The real part is visualized on the x-axis, which is now the real-numbers axis, and the y-axis is now the imaginary numbers axis. Addition and subtraction of complex numbers happen separately for each axis, just like the addition and subtraction of vectors.

$$z_1 = a_1 + ib_1, z_2 = a_2 + ib_2 z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)i \tag{11}$$

The *conjugate* of a complex number $z = a + bi$, denoted as z^* or $\bar{z}$, is $z^* = a - bi$. Make sure you understand how it looks like on the complex plane. The addition of $z + \bar{z}$ is $2\,\mathrm{Re}\{z\}$, while the subtraction $z - \bar{z} = 2\,\mathrm{Im}\{z\}$. This conjugation process mainly helps us when we have a complex fraction.

2.6 Polar Representation

When representing complex numbers in the complex plane, we can also look at their absolute value, and angle from the x-axis (Real axis). The length of the arrow in figure 3 is the absolute value of the complex number $|z| = \sqrt{a^2 + b^2} = r$. The angle is also called the phase or argument of the number, and it's received by $\varphi = \tan^{-1} {}^{b}\!/\!_{a}$. From simple trigonometrics we also find:

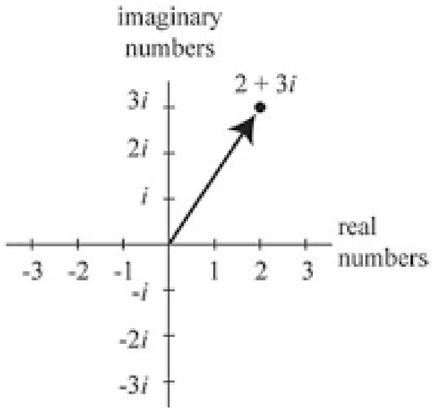

Figure 3: The complex plane.

$$\begin{aligned} r\cos\varphi &= a \\ r\sin\varphi &= b \\ z = r\cos\varphi + ir\sin\varphi &= r(\cos\varphi + i\sin\varphi) \\ &= re^{i\varphi} \end{aligned} \tag{12}$$

which is received from Euler's formula (Taylor series expansion of e).

2.7 Complex Roots

When we wish to take a root out of a complex number we would need to use the de Moivre's formula:

$$\sqrt[n]{z} = \sqrt[n]{r}e^{i\frac{\varphi+2\pi k}{n}} \tag{13}$$

2.8 Question

What is the answer for $z^3 = 4 - 4\sqrt{3}$?

2.9 answer

$$\begin{aligned} r &= \sqrt{4^2 + (4\sqrt{3})^2} = 8 \\ \varphi &= \tan^{-1}\frac{4}{4\sqrt{3}} = 300^\circ \\ z_k &= \sqrt[3]{8}e^{i\frac{300+2\pi k}{3}} = 2e^{i\frac{300+2\pi k}{3}} \end{aligned} \tag{14}$$

Part II
Linear Algebra

3 Definitions

3.1 Introduction

The basic motivation in linear algebra is to solve a system of linear equations such as this one:

$$\begin{aligned} a_{11}X + a_{12}Y + a_{13}Z &= b_1 \\ a_{21}X + a_{22}Y + a_{23}Z &= b_2 \\ a_{31}X + a_{32}Y + a_{33}Z &= b_3 \end{aligned} \tag{15}$$

Where a_{ij} are known parameters, X, Y, Z are the variables we wish to find and b_i are the solutions. What is the solution to this general problem?

First we'd like to find a more intuitive way to represent this problem. Generally speaking, there are three groups of mathematical entities in this group of equations: (1) The known parameters a_{ij}, the variables X, Y, Z and the known answers b_i. A possible way of writing it down could look like this:

$$a\vec{X} = \vec{b} \tag{16}$$

with the following notation:

$$\begin{aligned} a &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ \vec{X} &= \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \\ \vec{b} &= \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \end{aligned} \tag{17}$$

With this notation we can write equation 16 explicitly:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{18}$$

What exactly does it mean? Let's try to recreate the first line of equation 15: If we look closely we find it's identical to multiplying every entry in the first row of a with its corresponding variable in $\vec{X}$, and adding them all up. Thus we find that to receive the first entry in $\vec{b}$, which is b_1, we use the first row of a (but all three variables $\vec{X}$). Generally speaking:

$$a_{n1}X + a_{n2}Y + a_{n3}Z = b_n \tag{19}$$

Make sure you understand the reason behind both indices of a.

A matrix doesn't have to be symmetric. For example c, which is defined as

$$c = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{pmatrix} \tag{20}$$

is a 3 by 2 (3×2) matrix, with 3 rows and 2 columns, and behaves just like a when using it to solve linear equation systems. Technically speaking, we can also have a 1 by 1 matrix: $d = (d_{11})$, and the vectors we defined earlier are matrices as well. Since matrices are so broadly defined we can use them for multiple purposes, such as describing points in space, representing intensities in an image and so on. In order for us to do that, we have to define the algebra to which the matrices obey.

4 Algebra

4.1 Basic mathematical operations

4.1.1 Addition and subtraction

The definition of addition and subtraction is as simple as it gets - just add or subtract the corresponding entries,

$$\begin{aligned} A + B &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \pm \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & a_{13} \pm b_{13} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & a_{23} \pm b_{23} \end{pmatrix} \end{aligned} \tag{21}$$

just like regular numbers. Obviously, $A + B = B + A$ (commutative), which is not true for all matrix operations. Also, $A + (B + C) = (A + B) + C$ (associative property). Just to be clear, addition or subtraction of matrices of unequal dimensions in not defined.

4.1.2 Multiplying by a scalar

Next, we want to define how to multiply a "regular" number, which we call a scalar, with a matrix. The rule is simple - just multiply the scalar with each entry in the matrix:

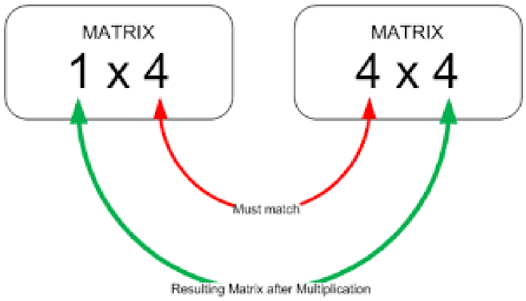

Figure 4: Matrix multiplication

$$a \cdot B = \begin{pmatrix} a \cdot b_{11} & a \cdot b_{12} \\ a \cdot b_{21} & a \cdot b_{22} \end{pmatrix} \tag{22}$$

Notice how B retains its dimensions after the multiplication.

4.2 Multiplying matrices

Now we wish to multiply two matrices. This definition seems more arbitrary than the previous one, and might be confusing for some. Before multiplying matrices, we have to make sure that the number of **columns** in the first matrix is equal to the number of **rows** in the second one (figure 4). Multiplication isn't defined without satisfying this condition. The final matrix will have the same number of rows as the first one, and the same number of columns as the second one.

The actual multiplication happens as follows: Take the **first row** of the first matrix, and multiply each element with its corresponding partner in the **first column** of the second matrix. Add all of the results - and that's the entry in the first row and first column of the solutions matrix:

$$\begin{aligned} A \cdot B &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} \end{pmatrix} \end{aligned} \tag{23}$$

Take note that generally $A \cdot B \neq B \cdot A$, and sometimes it's not even defined (when the dimensions are unequal). Matrix multiplication **is** associative, though, when it's defined.

4.3 Special Matrices

Identity Matrix We look for a matrix that is insensitive to multiplication with any other matrix, just like the number 1 is insensitive to multiplication with any number ($x \cdot 1 = 1 \cdot x = x$). We'll call it I, and one can easily show that it has to have this form:

$$I_3 \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{24}$$

where the subscript 3 notifies that its dimensions are 3 by 3 - as it's always a square matrix. A different way to indicate an identity matrix: $I_{ij} = \delta_{ij}$, where δ_{ij} is Kronecker's delta.

Zero Matrix The matrix that is insensitive to addition and subtraction is the zero matrix O, and we can easily see that it's just:

$$O_3 \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{25}$$

4.4 Questions

4.4.1 Question 1

Given the following matrix:

$$G = \begin{pmatrix} 0 & -2 & 1 & e \\ 24 & 4 & -3 & 2 \\ 1/5 & 0 & -\pi & 6 \end{pmatrix} \tag{26}$$

1. What's the value of G_{31}? *(Answer: 1/5)*

2. Can G be used to solve a system of linear equations? *(Answer: G only has 3 rows, which correspond to 3 equations - but 4 columns, which correspond to 4 unknown variables. Since we have less equations than variables, the answer is no.*

4.4.2 Question 2

$$\begin{aligned} AB &= \begin{pmatrix} 3 & 2 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & -2 \\ 3 & 3 & -1 \end{pmatrix} = ? \\ &= \begin{pmatrix} 9 & 6 & -8 \\ 11 & 12 & -2 \end{pmatrix} \end{aligned} \tag{27}$$

4.4.3 Question 3

Write down as many equivalent expressions as you can to:

$$\begin{aligned} A(B+C) &= ? \\ &= AB + AC \\ &= A(C+B) \\ &\neq (B+C)A \end{aligned} \tag{28}$$

5 Special matrices and matrix operators

5.1 Transpose

Transposing a matrix is switching between its rows and columns:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix} \tag{29}$$

or using the index notation we saw earlier:

$$(A^T)_{ij} = A_{ji} \tag{30}$$

Notice that if A is an $m \times n$ matrix, $A^T$is an $n \times m$ matrix. For example:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \tag{31}$$

Transposing is way to mathematically replace a column vector with a row vector, and vice versa:

$$\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \Rightarrow \vec{v}^T = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \tag{32}$$

In order to make sure that we can always multiply any given matrix (and vector) by itself ($A^2 = AA$) the convention is to use the transposed matrix in one of the entries, otherwise non-square matrices couldn't have been squared:

$$A^2 \equiv AA^T = A^T A \tag{33}$$

An important transpose property of the transpose operator for matrix Multiplication is:

$$(AB)^T = B^T A^T \tag{34}$$

And lastly

$$(A^T)^T = A \tag{35}$$

5.1.1 Question 4

Multiply the following vectors:

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix} \tag{36}$$

$$\vec{v}_1^T \vec{v}_2 = ? \; (= 1)$$

$$\vec{v}_1 \vec{v}_2^T = ? \begin{pmatrix} -1 & -2 & 0 \\ 1 & 2 & 0 \\ -3 & -6 & 0 \end{pmatrix}$$

5.2 Square and diagonal matrix

A square matrix is a matrix that has the same number of rows and columns ($n \times n$).

A diagonal matrix is a square matrix that only has non-zero entries in its diagonal:

$$D = \begin{pmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{pmatrix} \tag{37}$$

or again, in index notation: $D_{ij} = 0$ whenever $i \neq j$.

5.3 Symmetric and anti-symmetric matrix

A symmetric matrix is one that remains unchanged under transpose:

$$S^T = S \,(\Longleftrightarrow s_{ij} = s_{ji}) \tag{38}$$

An anti-symmetric matrix is one that changes its sign under the transpose operator:

$$A^T = -A \,(\Longleftrightarrow a_{ij} = -a_{ji}) \tag{39}$$

Comments:

- Both types of matrices have to be square.
- An anti-symmetric matrix has to have zeros on its diagonal.
- By definition, $(a+b)^T = a^T + b^T$
- For a general matrix a :
  - $S = a + a^T$ is symmetric, since $S^T = a^T + a = a + a^T = S$
  - $A = a - a^T$ is anti-symmetric, since $A^T = a^T - a = -(a - a^T) = -A$

5.4 Trace

For a square matrix a we'll define its trace as the sum of the diagonal entries:

$$Tr(A) = \sum_{i=1}^{n} A_{ii} \tag{40}$$

5.5 Determinant

The determinant is only defined for square matrices, and it can be thought of the scaling factor that this matrix represents. For a 2 by 2 matrix, its value is:

$$\det(A) = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{41}$$

5.6 Inverse

The inverse of a square matrix A, denoted as A^{-1}, exists there's a matrix B with the same dimensions such that:

$$AB = BA = I_n \tag{42}$$

where n is the matrix's dimension. If A cannot be inverted, it's called singular.

Part III
Probability and Statistics

6 Statistics

6.1 General statistical terms

6.1.1 Average

Average is the colloquial term for a the arithmetic mean. We tend to use "mean" in statistics and probability.

6.1.2 Mean

The central tendency of the dataset. When we say mean, we actually say "arithmetic mean", which is the sum of all items divided by the number of items (n):

$$\text{mean} = \bar{x} = \mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{43}$$

6.1.3 Median

The central number after sorting the set of numbers. In other words, a number that has the same amount of entries before and after it. When dealing with a set with an even number of items, we take the arithmetic mean of the two middle-ones, in which case the median won't be a part of the set. For a set with an odd number of items, its median will always be a part of that set. For example:

$$S = \{1, 2, 2, 2, 2, 2, 3, 199\} \Rightarrow \text{median}_S = 2 \tag{44}$$

As we can see from this contrived example, the median is less affected by outliers, while the mean will reflect the outlier more dramatically.

6.1.4 Mode

The most common number in a given set. When we have two values that are the most frequent, the set is called bimodal.

6.2 Standard deviation and variance

Standard deviation (SD) and variance are two statistical moments that tell us the dispersion, or spread, of the data. The variance (σ^2) is the average of the squared distance of each data point from the arithmetic mean:

$$Var(X) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{45}$$

The SD is the squared root of the variance, or just σ. Pay attention that only the SD has the same units as the original dataset. Moreover, this variance is called variance of population. The variance of a sample is similar, with the only change being that the N in the denominator turns to $N - 1$. This is because a sample is only a small part of a larger array of elements, and so the real variance of that larger array wasn't really measured. This $N - 1$ term makes it *unbiased*.

6.2.1 Properties of SD and variance

- $Var(X) \geq 0$
- $Var(X + a) = Var(X)$; $\sigma(X + a) = \sigma(X)$, where a is a constant.
- $Var(aX) = a^2Var(X)$; $\sigma(aX) = |a|\sigma(X)$

7 Questions

7.1 Question 1

Prove that $Var(aX) = a^2Var(X)$.

$$
\begin{aligned}
aX = a \cdot x_i = ax_i \; ; \; \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \rightarrow aX = a\bar{x} \\
Var(aX) &= \frac{1}{n}\sum_{i=1}^{n}(ax_i - a\bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} a^2(x_i - \bar{x})^2 \\
&= a^2\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= a^2 Var(X)
\end{aligned}
\tag{46}
$$

7.2 Random variables

7.2.1 Definition and examples

A random variable is a way to map the outcomes of a random process into numbers. For example, mapping a random coin flip to $X = 1$ if it's heads, and $X = 0$ if it's tails. One more example - The number of people inside a train station. This variable can obviously take a very large array of numbers, with different probabilities.

These were examples for a discrete random variable, but we can also think of a continuous random variable. For instance, the time it took a 100m runner to complete the run. Assuming our clock is perfect, we can think of endless possibilities for that time, meaning that our variable is now essentially continuous.

7.2.2 Probability distribution of a random variable

Creating a distribution of a discrete random variable is straightforward - as seen in figure 5. The sum of the y-axis values has to be 1. For a continuous variable we use a probability density function (PDF, figure 6) to describe the chance to receive each value. Here we find a subtlety, as the probability that the 100m race will be over after **exactly** 9.60 seconds is 0. For that reason we have to integrate a finite length of that curve in order to find the chance that the race will be over after 9.59-9.61 seconds. Similarly to the discrete case, here we have to make sure that the integral over the whole possible range of values is 1.

7.2.3 Cumulative distribution function

Another important function is the cumulative distribution function, or CDF. It represents the chance that the value of a random variable X will be less than or

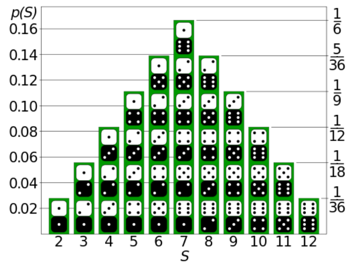

Figure 5: The probability mass function of the sum of counts from two dice. From Wikipedia.

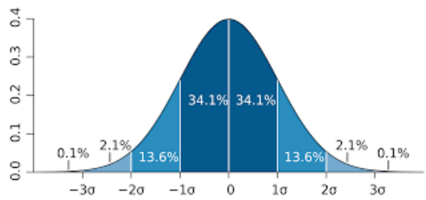

Figure 6: The probability density function of a random continuous variable. The percentage values indicate the chance of the value of the random variable to be inside that area. From Wikipedia.

equal to x:

$$F(x) = P(X \leq x)\ ;\ P(a < X \leq b) = F(b) - F(a) \tag{47}$$

For continuous random variables, integrating the probability density function results in the CDF:

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{48}$$

7.2.4 Expected value

The expected value of a random variable is the long-run average value of that variable, meaning after many repetitions of the experiment. For the discrete case:

$$E[X] = \frac{x_1 p_1 + x_2 p_2 + \ldots + x_n p_n}{p_1 + p_2 + \ldots + p_n} = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n \tag{49}$$

where p_1 is the probability to receive x_1 in our experiment, and n is the total number of possible values our random variable can get. For a continuous variable:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \tag{50}$$

where $f(x)$ is the PDF of X. Properties:

- $E[X + c] = E[X] + c$
- $E[X + Y] = E[X] + E[Y]$
- $E[aX] = aE[X]$
- $E[X] = \mu_x$ where μ is the mean of this variable.
- $Var(X) = E[(X - \mu_x)^2] = \sigma_x^2$
- All properties of the subtraction and addition of expected values of random variables are similar to mean and variance properties we discussed in random variables are similar to mean and variance properties we discussed in part 6.

7.3 Types of distributions of random variables

7.3.1 Binomial distribution - $B(n, p)$

The binomial distribution is the probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p. The chance to receive exactly $X = k$ successes (for example - 5 coin flips that result in exactly $k = 3$ heads) is:

$$Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \, ; \binom{n}{k} \equiv \frac{n!}{k!\,(n - k)!} \tag{51}$$

The expected value of a binomially-distributed random variable is $E[X] = np$, and the variance is $Var[X] = np(1 - p)$. When n is large, and p is far enough away from the extremes (0 or 1) one can estimate this distribution with a normal approximation $\mathcal{N}(np, np(1 - p))$ (which we'll discuss soon enough).

In addition, the binomial distribution converges towards the soon-to-be-defined Poisson distribution as the number of trials tends to infinity while the product np remains fixed, which is similar to saying that $p \to 0$. Note that the coefficient of the Poisson distribution is $\lambda = np$.

7.3.2 Poisson distribution - $P(\lambda)$

The Poisson distribution describes the probability that λ events will happen in a fixed interval. For example - the number of cars that pass in a junction in an hour is a random variable with a Poisson distribution. Here we assume that all hours are the same in regards to the probability of cars arriving there. We also assume that the fact that many cars arrived in this particular hour doesn't influence the number of cars in the next hour.

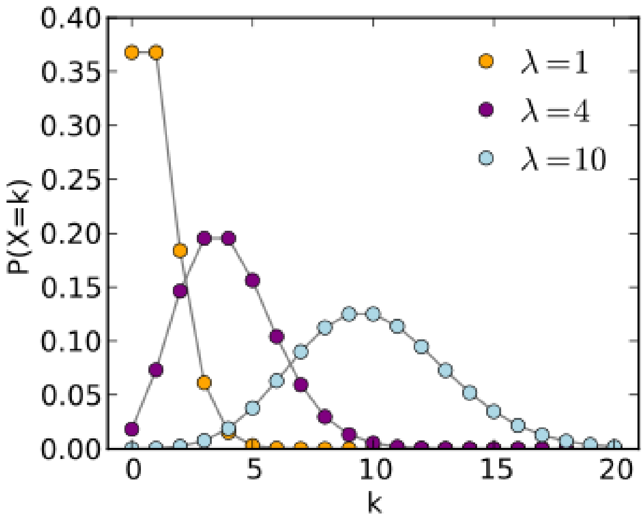

Figure 7: Probability mass function (PMF) of the Poisson process for different λ. From Wikipedia.

The expected value is $E(X) = \lambda$, and so is the variance, and the probability of observing exactly k events in our time bin is:

$$Pr(X = k \text{ events in interval}) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{52}$$

this is actually the PMF of the Poisson distribution (figure 7), and it's proven by looking at the limit of the binomial distribution when $n \to \infty$.

7.3.3 Normal distribution - $\mathcal{N}(\mu, \sigma^2)$

The normal distribution, or the bell curve, is a continuous probability distribution that is usually used when describing a large number of samples of random variables independently sampled. The values of these samples will distribute normally around the mean value.

The mean of the distribution (i.e. its center) is μ, and the variance is σ^2, making the standard deviation σ. Both median and mode are simply μ. The PDF is:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{53}$$

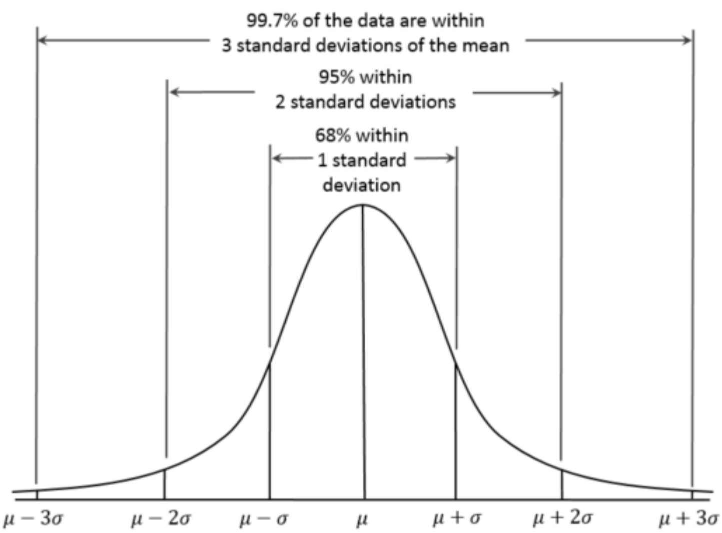

Figure 8: Normal distribution and the standard deviations around the mean. From Wikipedia.

8 Questions

8.1 Question 1 - from khanacademy.org

Steph makes 90% of the free-throws she attempts. Today she will shoot 3 free-throws. Assume that the results of free-throws are independent from each other. What is the probability that she makes exactly 2 of the 3 free throws?

Answer First we have to think which distribution is in play here. The right answer has to be binomial distribution, as we're talking about success and failure of trials.

With the binomial distribution as our guiding tool - we have a formula designed just to answer that question. But first let's think of the answer in a more intuitive manner. We've met the "Choose" operator $\binom{n}{k}$ which means, when dealing with binomial distributions, in how many ways can we arrange 2 scores in 3 attempts (disregarding order, no repetition):

$$\binom{3}{2} = \frac{3!}{2!\,(3-2!\,)} = \frac{6}{2} = 3 \tag{54}$$

Now lets look at one go, in which Steph scored the first and the last shots, missing the second. The chances of that happening are

$$P(score, miss, score) = 0.9 \cdot 0.1 \cdot 0.9 = 0.081 \tag{55}$$

and since we know we have 3 possible ways to arrange it, the answer is just:

$$P(\text{scores 2 of 3 throws}) = 3 \cdot P(\text{score, miss, score}) = 3 \cdot 0.081 = 0.243 \tag{56}$$

If we want to do it the formal way, we receive:

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \text{ ; } n = 3 \text{ tries ; } k = 2 \text{ successful shots} \\ P(X = 2) &= \binom{3}{2} \cdot 0.9^2 \cdot (1-0.9)^1 = 3 \cdot 0.81 \cdot 0.1 = 0.243 \end{aligned} \tag{57}$$