MSA Bootcamp

Aug 13-17
2018

Acknowledgements

Thanks to David Goldsman for providing me with LaTeX files for his lecture slides which comprise much of Part 1 and Part 3 of this document, as well as his ISyE 6739 Prob and Stat homework problems which are the the Workshop problems in Part 3.

David's GaTech website

`https://www2.isye.gatech.edu/people/faculty/David_Goldsman/`

Contents

1 Part 1: Integral Calculus (7 hours) **3**

- 1.1 Derivatives and Functions . . . 4
- 1.2 Newton's Method and Bisection . . . 7
- 1.3 **Derivative Problems** . . . 9
- 1.4 **Limit Problems** . . . 10
- 1.5 Workshop 1: Derivatives, functions, and limits. . . . 12
- 1.6 Integrals . . . 14
- 1.7 Integration by Parts and u-substitution Rules . . . 16
- 1.8 **IBP and u-sub Integration Problems** . . . 17
- 1.9 Trigonometric Integrals and Trig Substitution . . . 19
- 1.10 **Trig Integral and Trig Sub Problems** . . . 21
- 1.11 Partial Fractions . . . 22
- 1.12 Improper Integrals . . . 24
- 1.13 **Partial Frac and Improper Integrals Problems** . . . 25
- 1.14 Workshop 2: Integration . . . 28
- 1.15 Riemann Sums . . . 29
- 1.16 Taylor Series . . . 31
- 1.17 **Riemann Sum and Taylor Series Problems** . . . 33

2 Part 2: Linear Algebra (6.5 hours) **36**

- 2.1 Solving Systems of Linear Equations . . . 37
- 2.2 Determinants . . . 39
- 2.3 Workshop 3: Matrix algebra and linear system solving 40
- 2.4 Matrix Algebra . . . 43
- 2.5 Vector Spaces . . . 44
- 2.6 Workshop 4: Matrix algebra and vector spaces . . . 47
- 2.7 Linear Transformations . . . 50
- 2.8 **Linear Transformation Problems** . . . 51
- 2.9 Eigenvectors and Eigenvalues . . . 53
- 2.10 Diagonalization . . . 53
- 2.11 Workshop 5: Eigenvectors and eigenvalues . . . 55

3 Part 3: Probability and Statistics (4 hours) **57**
3.1 Basics: Conditional Probability . . . 57
3.2 Basics: Random Variables . . . 58
3.3 Well Known Discrete Distributions . . . 59
3.4 Continuous Random Variables and cdf's . . . 63
3.5 Well Known Continuous Distributions . . . 64
3.6 Simulating Simple Random Variables . . . 66
3.7 Expected Value . . . 67
3.8 Workshop 6: Random variables . . . 69
3.9 Moments and Standard Deviation . . . 78
3.10 Moment Generating Functions . . . 79
3.11 Functions of Random Variables . . . 80
3.12 Jointly Distributed Random Variables . . . 81
3.13 Independent Random Variables . . . 82
3.14 Covariance and Correlation . . . 84
3.15 Summary of Distributions . . . 85
3.16 Limit Theorems . . . 89
3.17 Statistics and Estimation . . . 90
3.18 Unbiased Estimation . . . 90
3.19 Distributional Results and Confidence Intervals . . . 91
3.20 Workshop 7: Advanced probability and statistics . . . 93

Chapter 1

Part 1: Integral Calculus (7 hours)

The purpose of the Bootcamp is to provide an accelerated lecture and workshop which will help you refresh the parts of mathematics that you once knew but may have forgotten, but will be very relevant for you throughout your MSA program. This document includes lecture notes and workshop problems for you to try yourself. It is *very* important that you try as many problems as you can, as the goal of the workshop is to make sure you are able to do similar problems *unassisted* during the MSA program. Everyone is expected to be at different levels coming in to the Bootcamp, so if you have any questions at any time please don't hesitate to ask. Each day will be split roughly in 3 parts (45 min lecture followed by a 15 min workshop, and then repeat).

There are three Parts to the MSA Bootcamp (each taking approx 1.5 days).

* Part 1: Integral Calculus, where derivatives of functions, integrals and area between curves, approximation methods, and Taylor series are discussed.

* Part 2: Linear Algebra, where we will discuss vector spaces and associated terminology: span, linear independence, basis; as well as the basics of solving linear systems of equations: Gaussian elimination, determinants, and matrix operations; and finally the theory of eigenvalues and eigenvectors will be discussed, including how to find them, as well as the associated concept of diagonalization.

* Part 3: Probability and Statistics, where we will discuss basic counting principles including the binomial theorem, nCr and nPr, basic probability theory over discrete and continuous variables, as well as limit theorems and hypothesis testing.

1.1 Derivatives and Functions

Definition We say that $f(x)$ is a *continuous* function if, for any x_0 and $x \in X$, we have $\lim_{x\to x_0} f(x) = f(x_0)$, where "lim" denotes a *limit* and $f(x)$ is assumed to exist for all $x \in X$.

Note: Another definition, the *formal definition of continuity*, you may have seen is the following: A function $f(x)$ is continuous at a if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \varepsilon$. This is written in mathematical notation with the universal quantifier $\forall$ (for all) and $\exists$ (there exists) as follows.

$$\forall \varepsilon > 0 \; \exists \delta > 0 \text{ s.t. } |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$$

Note: The notation $\lim_{x\to x_0}$ means "the limit as x tends to x_0" and is a *two-sided limit*. Meaning that $\lim_{x\to x_0^+} f(x) = \lim_{x\to x_0^-} f(x)$ the right-sided limit (from the + positive side) and the left-sided limit (from the - side) must agree. The limit $\lim_{x\to x_0} g(x) = L$ *exists* if whenever x is very close to x_0, the value of $g(x)$ is very close to L, $\forall \varepsilon > 0 \; \exists \delta > 0$ s.t. $|x - x_0| < \delta \implies |f(x) - L| < \varepsilon$.

Note: In gradeschool, we learned that to be *continuous* meant that you could draw the graph without lifting your pencil from the page, but a more sophisticated way of thinking about the definition is that a very small change in the x-value should make a very small change in the y-value. Think about both of these intuitive definitions in the following examples.

Example: The function $f(x) = 3x^2$ is continuous for all x. The function $f(x) = \lfloor x \rfloor$ (round down to the nearest integer, e.g., $\lfloor 3.4 \rfloor = 3$) has a "jump" discontinuity at any integer x.

□

Definition If $f(x)$ is continuous, then it is *differentiable* (has a *derivative*) if

$$f'(x) \;:=\; \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists and is well-defined for any given x. Think of the derivative as the slope of the function.

Note: Other notations for $f'(x)$ are $\frac{d}{dx}f(x)$ and $\frac{df}{dx}$, when the variable x is being emphasized.

Some well-known derivatives are:

$$\begin{aligned}
[x^k]' &= kx^{k-1}, \\
[e^x]' &= e^x, \\
[\sin(x)]' &= \cos(x), \\
[\cos(x)]' &= -\sin(x), \\
[\tan(x)]' &= \sec^2(x), \\
[\sec(x)]' &= \sec(x)\tan(x), \\
[\ln(x)]' &= \frac{1}{x}, \\
[\arctan(x)]' &= \frac{1}{1+x^2}.
\end{aligned}$$

Theorem Some well-known properties of derivatives are:

$$\begin{aligned}
[af(x)+b]' &= af'(x), \\
[f(x)+g(x)]' &= f'(x)+g'(x),
\end{aligned}$$

$$[f(x)g(x)]' = f'(x)g(x)+f(x)g'(x) \quad \text{(product rule)},$$

$$\left[\frac{f(x)}{g(x)}\right]' = \frac{g(x)f'(x)-f(x)g'(x)}{g^2(x)} \quad \text{(quotient rule)}^1,$$

$$[f(g(x))]' = f'(g(x))g'(x) \quad \text{(chain rule)}^2.$$

Example: Suppose that $f(x)=x^2$ and $g(x)=\ell\text{n}(x)$. Then

$$\begin{aligned}
[f(x)g(x)]' &= \frac{d}{dx}x^2\ell\text{n}(x) = 2x\ell\text{n}(x)+x, \\
\left[\frac{f(x)}{g(x)}\right]' &= \frac{d}{dx}\frac{x^2}{\ell\text{n}(x)} = \frac{2x\ell\text{n}(x)-x}{\ell\text{n}^2(x)}, \\
[f(g(x))]' &= f'(g(x))g'(x) = 2g(x)g'(x) = \frac{2\ell\text{n}(x)}{x}. \quad \square
\end{aligned}$$

Example: We can derive the formula for $\frac{d}{dx}\tan(x)$ using the quotient rule as

[1] Lo dee Hi minus Hi dee Lo over Lo Lo.
[2] www.youtube.com/watch?v=gGAiW5dOnKo

follows.

$$\begin{aligned}
\frac{d}{dx}\tan(x) &= \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) \\
&= \frac{(\cos x)(\sin x)' - (\sin x)(\cos x)'}{\cos^2 x} \quad \text{(quotient rule)} \\
&= \frac{(\cos x)^2 - (\sin x)(-\sin x)}{\cos^2 x} \\
&= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} \\
&= \frac{1}{\cos^2 x} \quad \text{(Pythagorus)} \\
&= \sec^2 x
\end{aligned}$$

□

Note: If $f'(x) > 0$ on the interval (a, b), then we say that f is *increasing* over (a, b). If $f'(x) < 0$ on the interval (a, b), then we say that f is *decreasing* over (a, b). Functions that are either non-increasing $f'(x) \leq 0$ or non-decreasing $f'(x) \geq 0$ over an interval are said to be *monotone*.

Note: The second derivative $f''(x) \equiv \frac{d}{dx}f'(x)$ and is the "slope of the slope." If $f(x)$ is "position," then $f'(x)$ can be regarded as "velocity," and as $f''(x)$ as "acceleration."

Example: The function $y = x^2 - 4x + 3 = (x-1)(x-3)$ is increasing over $(2, \infty)$, and decreasing over $(-\infty, 2)$. □

Definition The *critical points* of a function $y = f(x)$ are the x-values for which $f'(x) = 0$.

Note: The minimum or maximum of $f(x)$ can only occur at critical points, when the slope of $f(x)$ is zero, i.e., only when $f'(x) = 0$, say at $x = x_0$.

Definition If $f''(x) > 0$ for $x \in (a, b)$ then we say that f is *concave up* over (a, b), and if $f''(x) < 0$ for $x \in (a, b)$ we say that f is *concave down* over (a, b).

Note: Suppose x_0 is a critical point of f. Then if f is concave down at x_0, so that $f''(x_0) < 0$, then there is a maximum at x_0; if $f''(x_0) > 0$ so f is concave up, then you get a min; and if $f''(x_0) = 0$, you get a *point of inflection* provided the sign of f'' changes over x_0, from negative to positive or from positive to negative, so the concavity of f has to change over x_0 to have an inflection point (this is sometimes different for different professors/disciplines, in that some require $f''(x_0) = 0$ but not that the concavity of f to change over x_0).

Note: Concave up looks like a smile, and concave down looks like a frown. For example, $f(x) = (1 - x)(1 + x) = 1 - x^2$ is concave down on $(-\infty, \infty)$.

Example: The familiar function $y = x^2$ has only one critical point at $x = 0$. Since $f''(x) = 2 > 0$ for all x, the function is concave up everywhere and so any critical point, hence at $x = 0$ for instance, there is a minimum. $\square$

Example: Find the value of x that minimizes $f(x) = e^{2x} + e^{-x}$. The minimum can only occur when $f'(x) = 2e^{2x} - e^{-x} = 0$. After a little algebra, we find that this occurs at $x_0 = -(1/3)\ell n(2) \approx -0.231$. It's also easy to show that $f''(x) > 0$ for all x; and so x_0 yields a minimum. $\square$

1.2 Newton's Method and Bisection

Bisection: Suppose you can find x_1 and x_2 such that $g(x_1) < 0$ and $g(x_2) > 0$. (We'll follow similar logic if the inequalities are both reversed.) By the Intermediate Value Theorem (which you may remember), there must be a zero in $[x_1, x_2]$, that is, $x^\star \in [x_1, x_2]$ such that $g(x^\star) = 0$.

Thus, take $x_3 = (x_1 + x_2)/2$. If $g(x_3) < 0$, then there must be a zero in $[x_3, x_2]$. Otherwise, if $g(x_3) > 0$, then there must be a zero in $[x_1, x_3]$. In either case, you've reduced the length of the search interval by half.

Continue in this same manner until the length of the search interval is as small as desired.

Exercise: Try out the bisection method for $g(x) = x^2 - 2$, and come up with an interval approximation for $\sqrt{2}$, starting with $x_1 = 0$ and $x_2 = 2$.

Newton's Method: Suppose you can find a reasonable first guess for the zero, say, x_i, where we start off at iteration $i = 0$. If $g(x)$ has a nice, well-behaved derivative (which doesn't happen to be too flat near the zero of $g(x)$), then iterate your guess as follows:

$$x_{i+1} \;=\; x_i - \frac{g(x_i)}{g'(x_i)}.$$

Keep going until things appear to converge, for example that the decimal expansion of the numbers you calculate start stabilizing.

This makes sense since for x_i and x_{i+1} close to each other and the zero $x^\star$, we have

$$g'(x_i) \;\approx\; \frac{g(x^\star) - g(x_i)}{x^\star - x_i}.$$

Example: Use Newton's Method to find the root of $g(x) = x^2 - 2$, noting that the iteration step is to set

$$x_{i+1} \;=\; x_i - \frac{x_i^2 - 2}{2x_i} \;=\; \frac{x_i}{2} + \frac{1}{x_i}.$$

Let's start with a bad guess of $x_1 = 1$. Then

$$\begin{aligned} x_2 &= \frac{x_1}{2} + \frac{1}{x_1} = \frac{1}{2} + 1 = 1.5 \\ x_3 &= \frac{x_2}{2} + \frac{1}{x_2} \approx \frac{1.5}{2} + \frac{1}{1.5} = 1.4167 \\ x_4 &= \frac{x_3}{2} + \frac{1}{x_3} \approx 1.4142 \quad \text{Wow!} \quad \square \end{aligned}$$

The final method we will discuss right now, the linerization method, is just to use the derivative at a point $x = a$, which is just the slope of the tangent line at $(a, f(a))$, and a small step Δx, in order to approximate $f(x)$ for x-values that are near $x \approx a$. So if $\Delta x = x - a$ then

$$f(x) \approx f(a) + f'(a)\Delta x$$

Definition The *linearization* of a function $y = f(x)$ at $x = a$ is given by the formula $L(x) = f(a) + f'(a)(x - a)$. The linearization $y = L(x)$ is the tangent line of $y = f(x)$ at $x = a$, and gives a good approximation for the function $L(x) \approx f(x)$ near $x \approx a$.

Example: The linearization of $f(x) = \sin(x)$ at $x = 0$ is just

$$L(x) = \sin(0) + \cos(0)(x - 0) = x.$$

So $\sin(x) \approx x$ for values of x that are near $x \approx 0$. This is an informal way of justifying $\lim_{x\to 0} \frac{\sin(x)}{x} = 0$.

$\square$

Example: The linearization of $f(x) = \sqrt{x}$ at $x = 16$ is

$$L(x) = \sqrt{16} + \frac{1}{2\sqrt{16}}(x - 16).$$

So $\sqrt{x} \approx 4 + \frac{1}{8}(x - 16)$ for x-values near $x = 16$. So for example $\sqrt{17} \approx 4 + \frac{1}{8}(17 - 16) = 4 + \frac{1}{8} = 4.125$.

$\square$

1.3 Derivative Problems

1. $f(x) = \sin(\tan^{-1}(\sqrt{x^3+5x-2}))$

Solution: $f'(x) = \cos(\tan^{-1}(\sqrt{x^3+5x-2})) \cdot \frac{1}{1+(x^3+5x-2)} \cdot \frac{3x^2+5}{2\sqrt{x^3+5x-2}}$

□

2. $g(x) = \frac{3x^{1/4}e^{1/x}}{(x^4-\frac{1}{3x})^5(3x^2+2)^4}$

Solution: $\ln[g(x)] = \ln 3 + \frac{1}{4}\ln x + \frac{1}{x} - 5\ln(x^4 - \frac{1}{3x}) - 4\ln(3x^2+2)$, so:

$$\frac{g'(x)}{g(x)} = \frac{1}{4x} - \frac{1}{x^2} - \frac{5(4x^3+\frac{1}{3x^2})}{x^4-\frac{1}{3x}} - \frac{4(6x)}{3x^2+2}.$$

Then:

$$g'(x) = \frac{3x^{1/4}e^{1/x}}{(x^4-\frac{1}{3x})^5(3x^2+2)^4}\left[\frac{1}{4x} - \frac{1}{x^2} - \frac{5(4x^3+\frac{1}{3x^2})}{x^4-\frac{1}{3x}} - \frac{24x}{3x^2+2}\right].$$

□

3. $h(x) = (\ln x)^x$

Solution:

$$\ln[h(x)] = x\ln(\ln x)$$

So,

$$\frac{h'(x)}{h(x)} = \ln(\ln x) + x \cdot \frac{1}{\ln x} \cdot \frac{1}{x}.$$

Then:

$$h'(x) = (\ln x)^x\left[\ln(\ln x) + \frac{1}{\ln x}\right].$$

□

4. $k(x) = \log_2(\log_5(\log_6(8^{-3x})))$

Solution:

$$k'(x) = \frac{1}{\ln 2(\log_5(\log_6(8^{-3x})))} \cdot \frac{1}{\ln 5((\log_6(8^{-3x}))} \cdot \frac{1}{\ln 6(8^{-3x})} \cdot (-3)(\ln 8)(8^{-3x})$$

□

5. $s(t) = t^2\csc^3(5t)\sec^5(8t)$

Solution:

$$s'(t) = 2t\csc^3(5t)\sec^5(8t) - 15t^2\csc^3(5t)\cot(5t)\sec^5(8t)$$
$$+40t^2\csc^3(5t)\sec^5(8t)\tan(8t)$$

□

1.4 Limit Problems

1. Let $f(x) = \sqrt{5-x}$.

(a) Use the limit definition of the derivative to compute the derivative of the function.

(b) For what values of x is f differentiable? Write your answer as an interval.

Solution: :

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\
&= \lim_{h\to 0} \frac{1}{h}\left(\sqrt{5-x-h}-\sqrt{5-x}\right) \\
&= \lim_{h\to 0} \frac{1}{h}\left(\sqrt{5-x-h}-\sqrt{5-x}\right)\left(\frac{\sqrt{5-x-h}+\sqrt{5-x}}{\sqrt{5-x-h}+\sqrt{5-x}}\right) \\
&= \lim_{h\to 0} \frac{1}{h}\left(\frac{(5-x-h)-(5-x)}{\sqrt{5-x-h}+\sqrt{5-x}}\right) \\
&= \lim_{h\to 0} \left(\frac{-1}{\sqrt{5-x-h}+\sqrt{5-x}}\right) \\
&= \frac{-1}{2\sqrt{5-x}}
\end{aligned}$$

Differentiable on $(-\infty, 5)$.

□

2. Identify all points (x, y) on the graph of

$$g(x) = \frac{1}{3}x^3 - \frac{3}{2}x^2 + 1$$

where the tangent line is parallel to the line $8x - 2y = 1$.

Solution: : We want points on the curve where $g'(x)$ is equal to the slope of the line $8x - 2y = 1$.

$$\begin{aligned}
8x - 2y &= 1 \\
2y &= 8x - 1 \\
y &= 4x - 1/4
\end{aligned}$$

So the slope of the line is 4.

$$\begin{aligned}
4 &= g'(x) \\
4 &= \frac{d}{dx}\left(\frac{1}{3}x^3 - \frac{3}{2}x^2 + 1\right) \\
4 &= x^2 - 3x + 0 \\
0 &= x^2 - 3x - 4 \\
0 &= (x+1)(x-4)
\end{aligned}$$

At $x = -1$ and $x = 4$ the tangent line has the desired slope. Evaluating $g(x)$ at these points yields the points $(-1, -5/6)$, and $(4, -5/3)$.

□

3. Sketch a function, $y(x)$, that is defined on the domain $x \in [-4, 4]$, is continuous, odd, and not differentiable at exactly two points. Label your axes.

4. Give a formula for a function $y(x)$, that is continuous everywhere but not differentiable at $x = 1$. **Solution:** Many acceptable solutions, including $y(x) = |x - 1|$.

□

5. Compute the slope of the tangent line to $f(x)$ at the point where $x = 1$.

$$f(x) = \frac{5x+1}{4x^2+1}$$

Solution: :

$$\begin{aligned}
f'(x) &= \frac{d}{dx}\ \frac{5x+1}{4x^2+1} \\
&= \frac{\frac{d}{dx}(5x+1)(5x^2+1) - (5x+1)\frac{d}{dx}(4x^2+1)}{(4x^2+1)^2} \\
&= \frac{5(5x^2+1) - (5x+1)(8x)}{(4x^2+1)^2} \\
f'(1) &= \frac{5 \cdot 6 - 6 \cdot 8}{5^2} \\
&= \frac{30 - 48}{25} \\
&= -\frac{18}{25}
\end{aligned}$$

□

1.5 Workshop 1: Derivatives, functions, and limits.

1. Calculate the derivative using the definition of the derivative

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

(a) $f(x) = 4 - x^2$; $f'(-3), f`(0), f'(1)$

(b) $k(z) = \frac{1-z}{2z}$; $k'(1), k'(-1), k'(\sqrt{2})$

(c) $p(\theta) = \sqrt{3\theta}$; $p'(1), p'(2/3), p'(3)$

2. Check your answers from the previous problem by using the shortcut rules for calculating derivatives.

3. Find the slope of the tangent line at the given value of the independent variable.

(a) $f(x) = x + \frac{9}{x}, \quad x = -3$

(b) $k(x) = \frac{1}{2+x}, \quad x = 2$

4. Find the equation of the line $y = mx + b$ which is tangent to the curve $y = \frac{8}{\sqrt{x-2}}$ at the point $(6, 4)$.

5. Find the domain of $\sqrt{1 - x^2}$. Express your answer in interval notation.

6. Find the domain and range of $\ln(x - 1)$. Express your answer in interval notation.

7. For what x-values is the function $f(x)$ NOT continuous?

$$f(x) = \begin{cases} |x| & \text{if} \quad x \leq -1 \\ 2x + 1 & \text{if} \quad -1 < x < 2 \\ \frac{x-7}{x-3} & \text{if} \quad x \geq 2 \end{cases}$$

8. Let $f(x) = \sqrt{x - 1}$, and note that $\lim_{x \to 5} f(x) = 2$. Find the largest δ for which

$$|x - 5| < \delta \quad \Longrightarrow \quad |f(x) - 2| < 1.$$

9. For what value of a is the function $g(x)$ continuous at $x = 2$?

$$g(x) = \begin{cases} \dfrac{x^2 - x - 2}{x^2 - 4} & \text{if} \quad x \neq 2 \\ 3ax + 1 & \text{if} \quad x = 2 \end{cases}$$

10. Find $f'(x)$ using the definition of the derivative, where $f(x) = \frac{2}{x}$.

11. The derivative of $f(x)$ is $f'(x) = \frac{1}{2\sqrt{x}}$. What is the equation of the line tangent to the curve $y = f(x)$ at $x = 4$?

12. Suppose f, g, h are all functions from $\mathbb{R}$ to $\mathbb{R}$, and that $f(1) = f(2) = 3$, $g(1) = 2$, $g(3) = 4$, $h(2) = 1$, and $h(3) = 5$. Find the following:

(i) $f \circ h(2) =$

(ii) $g \circ f(1) =$

13. Find the limits.

(i) $\lim_{x \to 1^+} \dfrac{|1-x|}{x-1} =$

(ii) $\lim_{x \to \infty} \dfrac{x^3 - x^2 - x + 1}{3x^3 - 100} =$

(iii) $\lim_{x \to 3} \dfrac{x^2 - 9}{x^2 - 3x} =$

(iv) $\lim_{x \to 0} \dfrac{x^2 - 4}{x^2 - 1} =$

(v) For each function, identify the coordinates of any local extreme points and inflection points. Graph the function by finding the places where the function is increasing/decreasing/concave-up/concave-down.

(a) $y = 6 - 2x - x^2$
(b) $y = x(6 - 2x)^2$
(c) $y = \sin x \cos x$
(d) $y = \tan x$

(vi) Find the linearization of the function at the given x-value. Use your answer to approximate $f(x_0)$.

(a) $y = x^3$ at $x = 1$; $x_0 = 2.1$.
(b) $y = \sqrt{x}$ at $x = 4$; $x_0 = 5$.

1.6 Integrals

Definition The function $F(x)$ having derivative $f(x)$ is called the *antiderivative*. The antiderivative is denoted $F(x) = \int f(x)\,dx$; and this is also called the *indefinite integral* of $f(x)$.

Note: You can think of anti-derivatives as "undoing" what the derivative does. So $F'(x) = f(x)$ means $\int f(x)\ dx = F(x) + C$. We need to add the arbitrary constant "+C" because we lose any constant added to a function. For example, the functions $f(x) = x^2,\ g(x) = x^2 + 2,\ h(x) = x^2 - 100$ all have the same derivative $f'(x) = g'(x) = h'(x) = 2x$. So the antiderivative $\int 2x\ dx = x^2 + C$ which is in some sense *all of them.* Think of this as embodied in the statemtent: all the *vertical translates* of a function have the same slopes at every point.

We can thus rewrite our previous differentiation rules as integral rules, where just the left side and right side have been reversed essentially.

Example: Some well-known indefinite integrals are:

$$\begin{aligned}
\int x^k\,dx &= \frac{x^{k+1}}{k+1} + C \quad \text{for } k \neq -1 \\
\int \frac{dx}{x} &= \ell\text{n}|x| + C, \\
\int e^x\,dx &= e^x + C, \\
\int \cos(x)\,dx &= \sin(x) + C, \\
\int \frac{dx}{1+x^2} &= \arctan(x) + C,
\end{aligned}$$

where C is an arbitrary constant. □

Definition We define $\int_a^b f(t)\ dt$ to be the area under the curve $y = f(t)$ over the interval $[a, b]$, for a continuous function $f(t)$. We can compute this in a computer by doing Riemann sums by evaluating $y = f(t)$ at lots of points over the interval $[a, b]$ and then adding up the areas of the resulting rectangles. If we use a lot of rectangles we will get a very good approximation, and there are formulas which tell you (in any calculus textbook or just ask google for instance) how many rectangles you need depending on the size of the interval and some information about $y = f(t)$ (usually a bound on one of the higher order derivatives). For the sake of time we won't get into the computation of Riemann sums, but it is fairly straightforward. (Just google Riemann sums, or approximation formulas for Riemann sums if you are interested, or ask me during the workshop section)

The way we use integrals in calculus is usually one of two ways: either we want to know how an integral is changing, or we want to know how to compute a definite integral. These two ways are embodied in the FTC theorems below.

Fundamental Theorem of Calculus Part I: If $f(x)$ is continuous on the interval $[a, b]$, then the function $F(x) = \int_a^x f(t)\ dt$ is continuous on $[a, b]$ and differentiable on (a, b) and furthermore

$$F'(x) = \frac{d}{dx} \int_a^x f(t)\ dt = f(x).$$

Note: What this amazing theorem says is that the function $F(x) = \int_a^x f(t)\ dt$, which is the area under $y = f(t)$ over the interval $[a, x]$, has the property that if you look at the difference of

$$\int_a^x f(t)\ dt - \int_a^{x+h} f(t)\ dt$$

for very small values of h, and then divide by h (which is the length of the interval $[x, x + h]$ of course), then you get approximately $f(x)$, and in the limit you get *exactly* $f(x)$. I'll draw you a picture to help you see intuitively why this should be true if f is continuous.

The next FTC theorem tells you how to compute the value of an area under a curve.

Fundamental Theorem of Calculus Part II: If $f(x)$ is continuous with antiderivative $F(x)$, then the area under the curve for $x \in [a, b]$ is denoted and given by the *definite integral*.

$$\int_a^b f(x)\, dx \ = \ F(x)\Big|_a^b \ = \ F(b) - F(a).$$

Theorem Some well-known properties of definite integrals are:

$$\int_a^a f(x)\, dx \ = \ 0, \quad \text{(integrating over a point)}$$

$$\int_a^b f(x)\, dx \ = \ -\int_b^a f(x)\, dx, \quad \text{(integrating over a reversed interval)}$$

$$\int_a^b f(x)\, dx \ = \ \int_a^c f(x)\, dx + \int_c^b f(x)\, dx. \quad \text{(subdividing an interval)}$$

1.7 Integration by Parts and u-substitution Rules

Theorem Some other properties of general integrals are:

$$\int [cf(x) + g(x)]\, dx \;=\; c\int f(x)\, dx + \int g(x)\, dx, \quad \text{(linearity of integrals)}$$

$$\int f(x)g'(x)\, dx \;=\; f(x)g(x) - \int g(x)f'(x)\, dx \quad \text{(integration by parts)}^3,$$

$$\int u\ dv = uv - \int v\ du \quad \text{(alt formula for IBP)}$$

$$\int f(g(x))g'(x)\, dx \;=\; \int f(u)\, du \quad \text{(substitution rule)}^4.$$

Example: Using the substitution rule to integrate $\int_3^{e^2} \frac{1}{x\ln x}\ dx$, we set $u = \ln x$ so that $du = \frac{1}{x}\ dx$ and we can thus write the integral as

$$\begin{aligned}\int \frac{1}{x\ln x}\ dx &= \int \frac{1}{\ln x}\frac{1}{x}\ dx \\ &= \int \frac{1}{u}\ du \\ &= \ln(u) + C \\ &= \ln(\ln x) + C.\end{aligned}$$

So to evaluate the definite integral we can do it as follows

$$\begin{aligned}\int_3^{e^2} \frac{1}{x\ln x}\ dx &= \ln(u)\big|_{\ln 3}^{\ln e^2} \\ &= \ln(\ln x)\big|_3^{e^2} \\ &= \ln(\ln 3) - \ln(\ln e^2) \\ &= \ln(\ln 3) - \ln(2).\end{aligned}$$

□

Note: Notice that the limits of integration change if the independent variable is x, with x-values $x = 3, e^2$, or if the independent variable is $u = \ln x$, with *u-values* that are determined from the x-values and are $u = \ln 3, \ln(e^2)$.

Note: Note also that $\ln(e^2) = 2\ln(e) = 2$, from properties of $\ln(x)$.

[3] www.youtube.com/watch?v=OTzLVIc-O5E

[4] www.youtube.com/watch?v=eswQl-hcvU0

Example: Using integration by parts with $f(x) = x$ and $g'(x) = e^{2x}$ and the chain rule, we have

$$\int_0^1 xe^{2x}\,dx \;=\; \frac{xe^{2x}}{2}\bigg|_0^1 - \int_0^1 \frac{e^{2x}}{2}\,dx \;=\; \frac{e^2}{2} - \frac{e^{2x}}{4}\bigg|_0^1 \;=\; \frac{e^2+1}{4}. \quad \square$$

$\square$

Example: Integrating $\int x\sin x\;dx$ with $u = x$ and $dv = \sin x$, we have $du = dx$ and $v = -\cos(x)\;dx$, so $\int u\;dv = uv - \int v\;du$ becomes

$$\begin{aligned}\int x\sin x\;dx &= -x\cos x - \int -\cos x\;dx\\ &= -x\cos x + \int \cos x\;dx\\ &= -x\cos x + \sin x + C.\end{aligned}$$

$\square$

Note: Integration by parts is the *backwards product rule*, and the substitution rule is the *backwards chain rule*. So if you want to check your answer that $\int f(x)\;dx = F(x) + C$, then in order to check that $F'(x) = f(x)$: if you did IPB to find the integral $F(x)$, then you will use PROD to check the derivative $F'(x)$, and if you did u-sub to find the integral, then you will use CHAIN to check the derivative.

1.8 IBP and u-sub Integration Problems

Problem: Find the area bounded by the curves $f(x) = x^3 + 2x^2$ and $g(x) = x^2 + 2x$.

Solution: The curves intersect when $x = 0, 1, -2$. Breaking the interval into pieces, we see that f is larger on $[-2, 0]$ and g is larger on $[0, 1]$, so:

$$A = \int_{-2}^{0}[f(x) - g(x)]dx + \int_0^1 [g(x) - f(x)]dx.$$

Plugging in the functions and evaluating the integral yields a total area of $\frac{37}{12}$ square units.

$\square$

Problem: Evaluate the integrals:

1. $\int \frac{1}{x^2}\sec\left(\frac{1}{x}\right)\tan\left(\frac{1}{x}\right)dx$

Solution: Let $u = \frac{1}{x}$, then $du = -\frac{1}{x^2}dx$ and the integral becomes:

$$-\int \sec u \tan u du = -\sec u + C = -\sec\left(\frac{1}{x}\right) + C.$$

□

2. $\int x\tan^{-1}(x)dx$

Solution: Integration by parts: let $u = \tan^{-1}(x)$ and $dv = xdx$. Then $du = \frac{1}{1+x^2}dx$ and $v = \frac{x^2}{2}$, so

$$\int x\tan^{-1}(x)dx = \frac{x^2}{2}\tan^{-1}(x) - \int \frac{x^2}{2}\cdot\frac{1}{1+x^2}dx = \frac{x^2}{2}\tan^{-1}(x) - \frac{1}{2}\int \frac{x^2}{1+x^2}dx$$

$$= \frac{x^2}{2}\tan^{-1}(x) - \frac{1}{2}\int\left(1 - \frac{1}{1+x^2}\right)dx = \frac{x^2}{2}\tan^{-1}(x) - \frac{1}{2}\left[x - \tan^{-1}(x)\right] + C$$

$$= \frac{x^2}{2}\tan^{-1}(x) - \frac{x}{2} + \frac{1}{2}\tan^{-1}(x) + C.$$

□

Problem: Evaluate the following integrals:

1. $\int \frac{e^{2x}}{\sqrt{4-3e^{2x}}}dx$

Solution: Let $u = 4 - 3e^{2x}$, then $du = -6e^{2x}dx$ so $-\frac{1}{6}du = e^{2x}dx$. The integral becomes:

$$-\frac{1}{6}\int \frac{du}{\sqrt{u}} = -\frac{1}{3}\sqrt{u} + C = -\frac{1}{3}\sqrt{4-3e^{2x}} + C.$$

□

2. $\int_{-3}^{-2} \frac{dx}{\sqrt{4-(x+3)^2}}$

Solution: First, rewrite the integral. Pulling a 4 out of the denominator yields:

$$\frac{1}{2}\int_{-3}^{-2} \frac{dx}{\sqrt{1-\left(\frac{x+3}{2}\right)^2}}.$$

Now set $u = \frac{x+3}{2}$, then $du = \frac{1}{2}dx$. When $x = -3$, $u = 0$, and when $x = -2$, $u = \frac{1}{2}$, so the integral becomes:

$$\int_0^{\frac{1}{2}} \frac{du}{\sqrt{1-u^2}} = \sin^{-1}(u)|_0^{\frac{1}{2}} = \sin^{-1}\left(\frac{1}{2}\right) - \sin^{-1}(0) = \frac{\pi}{6} - 0 = \frac{\pi}{6}.$$

□

1.9 Trigonometric Integrals and Trig Substitution

Theorem From Pythagorus and simple algebraic manipulations we have

$$\begin{aligned} \sin^2(\theta) + \cos^2(\theta) &= 1 \\ \cos^2(\theta) &= 1 - \sin^2(\theta) \qquad (\text{EQ1}) \\ \sec^2(\theta) &= 1 + \tan^2(\theta) \qquad (\text{EQ2}) \\ \tan^2(\theta) &= \sec^2(\theta) - 1 \qquad (\text{EQ3}) \end{aligned}$$

We can use these identities to solve certain problems that have trigonometric functions in them by making an appropriate algebraic substitution followed by a u-substitution.

Example: In order to evaluate $\int \sin^2 x \cos^5 x\, dx$ first change two of the cosine's to sin's, and then make a u-sub as follows.

$$\begin{aligned} \int \sin^2 x \cos^5 x\, dx &= \int \sin^2 x \cos^4 x \cos x\, dx \\ &= \int \sin^2 x (\cos^2 x)^2 \;\cos x\, dx \\ &= \int \sin^2 x (1 - \sin^2 x)^2 \;\cos x\, dx \\ &= \int u^2 (1 - u^2)^2\, du \\ &= \int (u^2 - 2u^4 + u^6)\, du \\ &= \frac{1}{3}u^3 - \frac{2}{5}u^5 + \frac{1}{7}u^7 + C \\ &= \frac{1}{3}\sin^3 x - \frac{2}{5}\sin^5 x + \frac{1}{7}\sin^7 x + C. \end{aligned}$$

□

Example: In order to evaluate $\int \tan^3 x \sec^3 x\, dx$ first pull off one each of the $\tan x \sec x$ to become the du, with $u = \sec x$. Then change $\tan^2 x = \sec^2 x - 1$

to complete the set up for a u-substitution, as follows.

$$\begin{aligned}\int \tan^3 \sec^3 x\, dx &= \int \tan^2 x \sec^2 x\ \sec x \tan x\, dx \\ &= \int (\sec^2 x - 1)\sec^2 x\ \sec x \tan x\, dx \\ &= \int (u^2-1)u^2\, du \\ &= \frac{1}{5}u^5 - \frac{1}{3}u^3 + C \\ &= \frac{1}{5}\sec^5 x - \frac{1}{3}\sec^3 x + C \quad \square\end{aligned}$$

We can also use the identities to do trigonometric substitutions. This integral technique is probably one of the hardest to master, but the idea is to use the identities to make an integral that admits no obvious substitution into an integral that can be computed by one of the previously defined methods (IBP, u-sub, trig integral, or basic).

Example: In order to evaluate $\int \frac{1}{\sqrt{x^2-1}}\, dx$ we set $x = \sec\theta$ so that $x^2 - 1 = \sec^2\theta - 1 = \tan^2\theta$ using (EQ3). We then find $dx = \sec\theta\tan\theta\, d\theta$ in order to appropriately deal with the differentials in the integral. Thus, we have

$$\begin{aligned}\int \frac{1}{\sqrt{x^2-1}}\, dx &= \int \frac{1}{\sqrt{\sec^2\theta - 1}}\ \sec\theta\tan\theta\, d\theta \\ &= \int \frac{1}{\sqrt{\tan^2\theta}}\ \sec\theta\tan\theta\, d\theta \\ &= \int \sec\theta\, d\theta.\end{aligned}$$

Now, we use a clever algebraic manipulation and a u-substitution to solve the integral $\int \sec\theta\, d\theta$ as follows.

$$\begin{aligned}\int \sec\theta\, d\theta &= \int \sec\theta \frac{\sec\theta + \tan\theta}{\sec\theta + \tan\theta}\, d\theta \\ &= \int \frac{\sec^2\theta + \sec\theta\tan\theta}{\tan\theta + \sec\theta}\, d\theta \\ &= \ln(\tan\theta + \sec\theta) + C.\end{aligned}$$

Finally, we use the fact that

$$\sec\theta = \frac{hyp}{adj} \qquad \tan\theta = \frac{opp}{adj}$$

and pythagorus to solve for $\sec\theta$ and $\tan\theta$ in terms of x, using the definition that $\sec\theta = x = \frac{x}{1}$. In particular, $\tan\theta = \sqrt{x^2-1}$ and the integral becomes

$$\int \frac{1}{\sqrt{x^2-1}}\, dx = \ln(\sqrt{x^2-1} + x) + C.$$

1.10 Trig Integral and Trig Sub Problems

Problem: $\int \frac{dx}{x\sqrt{1+x^2}}$

Solution: Trig substitution: let $x = \tan\theta$, then $dx = \sec^2\theta d\theta$. So

$$\int \frac{dx}{x\sqrt{1+x^2}} = \int \frac{\sec^2\theta}{\tan\theta\sqrt{1+\tan^2\theta}}d\theta = \int \frac{\sec^2\theta}{\tan\theta\sqrt{\sec^2\theta}}d\theta = \int \frac{\sec^2\theta}{\tan\theta\sec\theta}d\theta = \int \frac{\sec\theta}{\tan\theta}d\theta$$

$$= \int \frac{1/\cos\theta}{\sin\theta/\cos\theta}d\theta = \int \frac{1}{\sin\theta}d\theta = \int \csc(\theta)d\theta = -\ln|\csc\theta + \cot\theta| + C$$

$$= -\ln\left|\frac{\sqrt{1+x^2}}{x} + \frac{1}{x}\right| + C.$$

□

Problem: $\int \sqrt{25-x^2}dx$

Solution: Trig substitution: let $x = 5\sin\theta$, then $dx = 5\cos\theta d\theta$. Then

$$\int \sqrt{25-x^2}dx = \int(\sqrt{25-25\sin^2\theta})(5\cos\theta)d\theta = \int(\sqrt{25\cos^2\theta})(5\cos\theta)d\theta$$

$$= \int(5\cos\theta)(5\cos\theta)d\theta = 25\int\cos^2\theta d\theta = \frac{25}{2}\int[1+\cos(2\theta)]d\theta = \frac{25}{2}\left[\theta + \frac{1}{2}\sin(2\theta)\right]+C$$

$$= \frac{25}{2}\left[\sin^{-1}\left(\frac{x}{5}\right) + \frac{x}{5}\cdot\frac{\sqrt{25-x^2}}{5}\right] + C = \frac{25}{2}\sin^{-1}\left(\frac{x}{5}\right) + \frac{x\sqrt{25-x^2}}{2} + C.$$

□

Problem: $\int \tan^3(x)\sec^4(x)dx$

Solution:

$$\int \tan^3(x)\sec^4(x)dx = \int \tan^3(x)\sec^2(x)\sec^2(x)dx = \int \tan^3(x)[1+\tan^2(x)]\sec^2(x)dx$$

$$= \int[\tan^3(x)+\tan^5(x)]\sec^2(x)dx = \int(u^3+u^5)du \quad [u = \tan(x)] = \frac{1}{4}\tan^4(x)+\frac{1}{6}\tan^6(x)+C.$$

□

1.11 Partial Fractions

We can use the partial fraction decomposition technique to solve integrals that do not admit other solutions. For example, since

$$\frac{1}{1-x^2} = \frac{1}{2}(\frac{1}{1-x} + \frac{1}{1+x}),$$

we can integrate

$$\int \frac{1}{1-x^2}\,dx = \frac{1}{2}\int\left(\frac{1}{1-x} + \frac{1}{1+x}\right)\,dx = \frac{1}{2}\ln(1-x) + \frac{1}{2}\ln(1+x) + C.$$

One way to obtain the partial fraction decomposition is to adhere to the following method.

Partial Fraction Decomposition:

1. Write the quotient with $g(x)$ factored into linear and irreducible quadratic factors:

$$g(x) = (x-r_1)^{n_1}\cdots(x-r_k)^{n_k}(x^2+p_1x+q_1)^{m_1}\cdots(x^2+p_\ell x+q_\ell)^{m_\ell}.$$

2. For each repeated linear factor of $g(x)$, expand

$$\frac{f(x)}{g(x)} = \cdots + \frac{A_1}{(x-r)} + \frac{A_2}{(x-r)^2} + \cdots + \frac{A_n}{(x-r)^n} + \cdots$$

3. For each repeated irreducible quadratic factor of $g(x)$, expand

$$\frac{f(x)}{g(x)} = \cdots + \frac{A_1x+B_1}{(x^2+px+q)} + \frac{A_2x+B_2}{(x^2+px+q)^2} + \cdots + \frac{A_mx+B_m}{(x^2+px+q)^m} + \cdots$$

4. Set the original fraction $f(x)/g(x)$ equal to the sum of all these partial fractions and clear fractions by finding a common denominator of the partial fractions, arranging the terms in decreasing powers of x.

5. Equate the coefficients of corresponding powers of x and solve the resulting equations for the undetermined coefficients.

The method is elucidated in the following examples.

Example: Using partial fractions to evaluate $\int \frac{x^2+4x+1}{(x-1)(x+1)(x+3)}\,dx$, we see that the denominator has three linear factors, none of which are repeated, and no irreducible quadratic factors. The partial fraction decomposition thus has the form

$$\frac{x^2+4x+1}{(x-1)(x+1)(x+3)} = \frac{A}{x-1} + \frac{B}{x+1} + \frac{C}{x+3}.$$

To find the values of the undetermined coefficients A, B, and C, we clear fractions and get

$$\begin{aligned} x^2+4x+1 &= A(x+1)(x+3)+B(x-1)(x+3)+C(x-1)(x+1) \\ &= A(x^2+4x+3)+B(x^2+2x-3)+C(x^2-1) \\ &= (A+B+C)x^2+(4A+2B)x+(3A-3B-C). \end{aligned}$$

The polynomials on both sides (the first polynomial and the last one above) are equal to each other, so we equate coefficients of like powers of x, obtaining the following three equalities

$$\begin{aligned} \text{Coefficient of } x^2: &\quad A+B+C=1 \\ \text{Coefficient of } x: &\quad 4A+2B \quad =4 \\ \text{Constant:} &\quad 3A-3B-C=1 \end{aligned}$$

There are several ways of solving such a system of linear equations for the unknowns A, B, and C (which we will get to in Part 2 tomorrow). Whatever method is used, the solution is $A = 3/4$, $B = 1/2$, and $C = -1/4$. Hence we have

$$\begin{aligned} \int \frac{x^2+4x+1}{(x-1)(x+1)(x+3)}\,dx &= \frac{(3/4)}{x-1}+\frac{(1/2)}{x+1}+\frac{(-1/4)}{x+3}\,dx \\ &= \frac{3}{4}\ln(x-1)+\frac{1}{2}\ln(x+1)-\frac{1}{4}\ln(x+3)+Constant. \end{aligned}$$

□

Example: Using partial fractions to evaluate $\int \frac{-2x+4}{(x^2+1)(x-1)^2}\,dx$ we have that the denominator has one irreducible quadratic (non-repeated) factor and a repeated linear factor. Thus the partial fraction decomposition of the integrand takes the form

$$\frac{-2x+4}{(x^2+1)(x-1)^2} = \frac{Ax+B}{x^2+1}+\frac{C}{(x-1)}+\frac{D}{(x-1)^2}.$$

Clearing the equation of fractions gives

$$\begin{aligned} -2x+4 &= (Ax+B)(x-1)^2+C(x-1)(x^2+1)+D(x^2+1) \\ &= (A+C)x^3+(-2A+B-C+D)x^2+(A-2B+C)x+(B-C+D). \end{aligned}$$

Equating coefficients of like terms gives

$$\begin{aligned} \text{Coefficient of } x^3: &\quad A+C \quad =0 \\ \text{Coefficient of } x^2: &\quad \text{-}2A+B-C+D=0 \\ \text{Coefficient of } x: &\quad A-2B+C=-2 \\ \text{Constant:} &\quad B-C+D=4 \end{aligned}$$

Solving to find the values of A, B, C, D we find that $A = 2, C = -2, B = 1, D = 1$. We substitute these values into our partial fraction decomposition and then integrate

$$\begin{aligned}
\int \frac{-2x+4}{(x^2+1)(x-1)^2}\,dx &= \int \left(\frac{2x+1}{x^2+1} - \frac{2}{(x-1)} + \frac{1}{(x-1)^2} \right) dx \\
&= \int \left(\frac{2x}{x^2+1} + \frac{1}{x^2+1} - \frac{2}{x-1} + \frac{1}{x-1}^2 \right) dx \\
&= \ln(x^2+1) + \tan^{-1} x - 2 - 2\ln(x-1) - \frac{1}{x-1} + C.
\end{aligned}$$

□

1.12 Improper Integrals

Integrals which have infinity in them, either as one of the limits of integration or for which the integrands themselves become infinite over the interval of integration, are called *improper integrals*.

The way to deal with improper integrals is to replace the ∞ symbol with a limit, in the first case, or to split up the integral and replace the x-value for which $f(x)$ becomes infinite with a limit, in the second case.

Example: The improper integral

$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{N\to\infty} \int_1^N \frac{1}{x^2}\,dx = \lim_{N\to\infty} \left. \frac{-1}{x} \right|_1^N = \lim_{N\to\infty} \left(\frac{-1}{N} + 1 \right) = 1.$$

So the area under $y = \frac{1}{x^2}$ over the interval $(1, \infty)$ is finite, and equals 1.

□

Example: The improper integral

$$\int_1^\infty \frac{1}{\sqrt{x}}\,dx = \lim_{N\to\infty} \int_1^N \frac{1}{\sqrt{x}}\,dx = \lim_{N\to\infty} 2\sqrt{x}\Big|_1^N = \lim_{N\to\infty} \left(2\sqrt{N} - 2 \right) = +\infty \text{ DNE}.$$

So the area under $y = \frac{1}{\sqrt{x}}$ over the interval $(1, \infty)$ is infinite, and does not exist.

□

Definition We say that an improper integral which is infinite (or does not exist) is *divergent* or that it *diverges*. We say that an improper integral which is finite (for which the limit exists) is *convergent* or that it *converges*.

Note: The improper integral $\int_1^\infty \frac{1}{x^p}\,dx$ converges if $p > 1$, and diverges if $0 < p \le 1$.

1.13 Partial Frac and Improper Integrals Problems

Problem: $\int \frac{x+1}{x^2(x-1)} dx$

Solution: Using partial fractions:

$$\frac{x+1}{x^2(x-1)} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x-1}.$$

Multiply both sides by the common denominator, and combine like terms to get

$$x+1 = (A+C)x^2 + (-A+B)x - B.$$

Solving for A, B, and C yields $A = -2$, $B = -1$, and $C = 2$. Then

$$\int \frac{x+1}{x^2(x-1)} dx = \int \left(\frac{-2}{x} + \frac{-1}{x^2} + \frac{2}{x-1} \right) dx = -2\ln|x| + \frac{1}{x} + 2\ln|x-1| + C.$$

□

Problem: $\int \frac{x+1}{x^2-4x+8} dx$

Solution: Since we cannot factor the denominator any further, we use the method for rational functions. First, we try to obtain the derivative of the denominator in the numerator, then we break up our problem into two separate integrals:

$$\int \frac{x+1}{x^2-4x+8} dx = \frac{1}{2} \int \frac{2x+2}{x^2-4x+8} dx = \frac{1}{2} \int \frac{2x-4+4+2}{x^2-4x+8} dx$$

$$= \frac{1}{2} \int \frac{2x-4}{x^2-4x+8} dx + 3 \int \frac{dx}{x^2-4x+8} = \frac{1}{2} \ln|x^2-4x+8| + 3 \int \frac{dx}{x^2-4x+4-4+8}$$

$$= \frac{1}{2} \ln|x^2-4x+8| + 3 \int \frac{dx}{(x-2)^2+4} = \frac{1}{2} \ln|x^2-4x+8| + \frac{3}{2} \tan^{-1}\left(\frac{x-2}{2} \right) + C.$$

□

Problem: $\int \frac{x+2}{x+1} dx$

Solution: Carrying out the long division, we have:

$$\int \frac{x+2}{x+1} dx = \int \left(1 + \frac{1}{x+1} \right) dx = x + \ln|x+1| + C.$$

□

Problem: Evaluate the improper integral if it converges, or show that the integral diverges.

$$\int_1^3 \frac{1}{(x^2-1)^{3/2}} dx$$

Solution: There is a vertical asymptote at $x = 1$. Rewrite the integral as follows:

$$\int_1^3 \frac{1}{(x^2-1)^{3/2}}dx = \lim_{a\to 1^+} \int_a^3 \frac{1}{(x^2-1)^{3/2}}dx.$$

We need to find a general antiderivative for the function $f(x) = \frac{1}{(x^2-1)^{3/2}}$. Using trig sub, set $x = \sec\theta$, then $dx = \sec\theta\tan\theta d\theta$ and:

$$\int \frac{1}{(x^2-1)^{3/2}}dx = \int \frac{\sec\theta\tan\theta d\theta}{(\sec^2\theta-1)^{3/2}} = \int \frac{\sec\theta\tan\theta d\theta}{\tan^3\theta} = \int \frac{\sec\theta}{\tan^2\theta}d\theta = \int \frac{\cos\theta}{\sin^2\theta}d\theta$$

$$= -\frac{1}{\sin\theta} + C = -\frac{x}{\sqrt{x^2-1}} + C.$$

So the improper integral becomes:

$$\int_1^3 \frac{1}{(x^2-1)^{3/2}}dx = \lim_{a\to 1^+}\left[-\frac{x}{\sqrt{x^2-1}}\right]_a^3 = \lim_{a\to 1^+}\left[-\frac{3}{\sqrt{8}} + \frac{a}{\sqrt{a^2-1}}\right] = \infty,$$

so the integral **diverges**.

□

Problem: Evaluate the improper integral if it converges, or show that the integral diverges.

$$\int_0^\infty x^2 e^{-2x}dx$$

Solution: Rewrite as follows, and use integration by parts (twice) or tabular integration to obtain the anti-derivative:

$$\int_0^\infty x^2e^{-2x}dx = \lim_{b\to\infty}\int_0^b x^2e^{-2x}dx = \lim_{b\to\infty}\left[-\frac{1}{2}x^2e^{-2x} - \frac{1}{2}xe^{-2x} - \frac{1}{4}e^{-2x}\right]_0^b$$

$$= \lim_{b\to\infty}\left[-\frac{1}{2}b^2e^{-2b} - \frac{1}{2}be^{-2b} - \frac{1}{4}e^{-2b} + 0 + 0 + \frac{1}{4}\right] = \frac{1}{4},$$

so the integral converges. (NOTE: you need L'Hopital's Rule to evaluate the limits $\lim_{b\to\infty} b^2e^{-2b}$ and $\lim_{b\to\infty} be^{-2b}$.)

□

Problem: For what values of p does the integral $\int_4^\infty \frac{dx}{x(\ln x)^p}$ converge?

Solution: Rewrite the integral as a limit:

$$\int_4^\infty \frac{dx}{x(\ln x)^p} = \lim_{b\to\infty}\int_4^b \frac{dx}{x(\ln x)^p}.$$

Let $u = \ln x$, then $du = \frac{1}{x}dx$. When $x = b$, $u = \ln b$, and when $x = 4$, $u = \ln 4$, so the integral becomes:

$$\lim_{b\to\infty}\int_{\ln 4}^{\ln b}\frac{du}{u^p}.$$

If $p \neq 1$, then:

$$\int \frac{du}{u^p} = \frac{u^{-p+1}}{-p+1} + C,$$

and if $p = 1$, then:

$$\int \frac{du}{u} = \ln|u| + C.$$

So:

$$\int_4^\infty \frac{dx}{x(\ln x)^p} = \lim_{b\to\infty} \frac{(\ln b)^{1-p}}{1-p} - \frac{(\ln 4)^{1-p}}{1-p}, p \neq 1$$

and

$$\int_4^\infty \frac{dx}{x(\ln x)^p} = \ln|\ln b| - \ln|\ln 4|, p = 1.$$

When $p = 1$, this limit is ∞ and the integral **diverges**.
When $p > 1$, $1 - p < 0$, so as $b \to \infty$, $\ln b \to \infty$ and thus $(\ln b)^{1-p} \to 0$, so the limit is $-\frac{(\ln 4)^{1-p}}{1-p} = \frac{1}{(\ln 4)^{p-1}(p-1)}$, a finite value, so the integral **converges**.
When $p < 1$, $1 - p > 0$, so as $b \to \infty$, $\ln b \to \infty$ and $\ln b^{1-p} \to \infty$, so the limit is ∞ and the integral **diverges**.
Thus, the integral only converges when $p > 1$.

□

Problem: Find the area bounded by the curve $y = \frac{1}{x^2+9}$, the x-axis, and $x \geq 0$.

Solution: We need to evaluate:

$$\int_0^\infty \frac{dx}{x^2+9} = \lim_{b\to\infty} \frac{1}{9} \int_0^b \frac{dx}{\left(\frac{x}{3}\right)^2 + 1} = \frac{1}{9} \lim_{b\to\infty} 3\tan^{-1}\left(\frac{x}{3}\right)|_0^b$$

$$= \frac{1}{3} \lim_{b\to\infty} \left[\tan^{-1}\left(\frac{b}{3}\right) - \tan^{-1} 0\right] = \frac{1}{3}\left[\frac{\pi}{2} - 0\right] = \frac{\pi}{6} \quad units^2.$$

□

1.14 Workshop 2: Integration

1. Solve the separable differential equation.

$$\frac{dy}{dx} = xe^{x/3}(4y^2+1).$$

2. **(a)** Find a closed formula for the n-th term of the sequence in terms of n. Then, **(b)** find the limit of the sequence using L'Hopital's rule.

$$\frac{3}{5}, \frac{-6}{11}, \frac{12}{17}, \frac{-24}{23}, \frac{48}{29}, \ldots$$

3. Integrate.

(a) $\displaystyle\int 5\sin^5(x/3)\cos^2(x/3)\,dx$

(b) $\displaystyle\int \frac{\sqrt{x^2-9}}{x}\,dx$

(c) $\displaystyle\int \frac{x+2}{x^3+x}\,dx$

4. Evaluate the limit using L'Hopital's rule.

$$\lim_{x\to\infty}(x+e^x)^{1/x}$$

5. Find the approximate area bounded by the curve $y = \sin(x)$ over the interval $[0, \pi/2]$ using the trapezoid rule with $n = 3$ trapezoids. For up to 4 points, check your answer with FTC.

6. Evaluate the improper integral.

$$\int_1^\infty \frac{1}{x(2+\ln x)^2}\,dx$$

7. Suppose a population of bacterium, left unchecked, follows the rule that the current rate of growth of the population is proportional to the current population size. If the bacterium colony starts with 8 members, and after 2 days there are 400 bacterium, then what is the function which gives the number of bacterium in the colony after t days?

Riemann Sums and Taylor Series

In this section we will use *sigma notation.* A sigma is Σ and it means "to sum". So if you see

$$\sum_{i=1}^{N} x_i$$

that just means

$$x_1 + x_2 + x_3 + \cdots + x_N.$$

Sometimes shorthand notation is used so you may see instead $\sum_i x_i$ or even just $\sum x_i$ where the index is just assumed to be explained in context.

We are going to be using sigma notation to describe two important concepts, and the last two topics in Part 1, which are Riemann Sums and Taylor Series.

1.15 Riemann Sums

Riemann Sums are used to approximate the area under a curve if you don't have the ability to compute exactly. For instance, the simplest is to approximate the area under a continuous function $f(x)$ from a to b by adding up the areas of n adjacent rectangles of width $\Delta x = (b-a)/n$ and height $f(x_k)$, where $x_k = a + k\Delta x$ is the right-hand endpoint of the kth rectangle. Thus,

$$\int_a^b f(x)\,dx \quad \approx \quad \sum_{k=1}^{n} f(x_k)\Delta x \;=\; \frac{b-a}{n}\sum_{i=k}^{n} f\left(a + \frac{k(b-a)}{n}\right).$$

In fact, as $n \to \infty$, this approximation becomes an equality, and its actually equivalent to define $\int_a^b f(x)\,dx$ as the limit of Riemann sums.

Example: To approximate the integral $\int_0^1 \sin(\pi x/2)\,dx$ with n rectangles of equal size we have $\Delta x = 1/n$ and $x_i = i/n$, which simplifies the notation a bit. Then

$$\begin{aligned}
\int_a^b f(x)\,dx &= \int_0^1 f(x)\,dx \\
&\approx \sum_{i=1}^{n} f(x_i)\Delta x \\
&= \frac{1}{n}\sum_{i=1}^{n} \sin\left(\frac{\pi i}{2n}\right).
\end{aligned}$$

For $n = 100$, this calculates out to a value of 0.6416, which is pretty close to the true answer of $2/\pi \approx 0.6366$. $\square$

In the last example, we used the right-endpoint of each rectangle to evaluate. But we could have used the left-endpoint, or the midpoint of each interval, or we could have used whichever endpoint gave us the larger value, or the smaller value.

Definition A *Riemann sum* is a sum of the form

$$\sum_{k=1}^{n} f(c_k)\Delta x_k$$

where $\{x_k|\ 1 \le i \le n\}$ form a partition of the interval $[a, b]$ over which $y = f(x)$ is continuous, $c_k \in [x_k, x_{k+1}]$ and $\Delta x_k = x_{k+1} - x_k$, for $1 \le k < n$.

The most common Riemann sums are the upper, lower, left-endpoint, and right-endpoint Riemann sums. For each of these, the points $x_k = a + \frac{k(b-a)}{n}$, or $x_k = a + k\Delta x$.

This just means that the interval $[a, b]$ is subdivided equally into n equal pieces, and the endpoints of these intervals of length $(b - a)/n$ are called x_k.

Note: Actually there's no rule that the interval $[a, b]$ has to be equally subdivided, all that is required to get convergence to the definite integral is that the maximum width of any sub-interval (the "mesh size") has to go to zero.

Remark Sometimes it is useful to find a closed formula for a Riemann sum. The idea is to find a formula R_n which calculates the approximate area under $f(x)$ over $[a, b]$ using n rectangles, but *doesn't require you to add n separately computed values.*

Example: For example, we can use the following two formulas to find a closed formula for the approximate area $\int_2^5 x^2\, dx$ using n rectangles.

$$\begin{aligned} \sum_{k=1}^{n} k &= \frac{n(n+1)}{2}, \\ \sum_{k=1}^{n} k^2 &= \frac{n(n+1)(2n+1)}{6} \end{aligned}$$

The Riemann sum we want to find a closed formula for is

$$\begin{aligned} R_n &= \sum_{k=1}^{n} \left(2 + \frac{k(5-2)}{n}\right)^2 \frac{(5-2)}{n} \\ &= \frac{3}{n}\sum_{k=1}^{n}(2 + \frac{3k}{n})^2 \\ &= \frac{3}{n}\sum_{k=1}^{n}(4 + \frac{12k}{n} + \frac{9k^2}{n^2}) \\ &= \frac{3}{n} \cdot 4\sum_{k=1}^{n} 1 + \frac{3}{n} \cdot \frac{12}{n}\sum_{k=1}^{n} k + \frac{3}{n} \cdot \frac{9}{n^2}\sum_{k=1}^{n} k^2 \\ &= \frac{12}{n} \cdot n + \frac{36}{n^2} \cdot \frac{n(n+1)}{2} + \frac{27}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} \end{aligned}$$

Which can be written after simplifying as

$$R_n = 12 + \frac{18(n+1)}{n} + \frac{9(n+1)(2n+1)}{2n^2}.$$

Note that $R_n \to 12 + 18 + 9 = 39$ as $n \to \infty$, and indeed $\int_2^5 x^2\ dx = \frac{1}{3}x^3|_2^5 = \frac{1}{3}(5^3 - 2^3) = \frac{125-8}{3} = \frac{117}{3} = 39$.

□

1.16 Taylor Series

Taylor series are probably the single most powerful way of looking at a function. For example, consider the Taylor series about $x = 0$ for the following three functions.

$$\begin{aligned} \sin(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}, \\ \cos(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}, \\ e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!}. \end{aligned}$$

Note: We use the notation $k!$ to mean $k! = k(k-1)(k-2)\cdots 3\cdot 2\cdot 1$, called "$k$ factorial". It is a very common notation in counting problems in mathematics, and in probability and statistics.

It turns out that each of the three series above converge for all values of x, but the convergence is very fast near $x = 0$. For example, taking only 3 terms on the right hand side (RHS), we have

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

Just ponder this for a minute.

I'll wait. No really go ahead.

What this means is that to a very high degree of accuracy, $\sin x$ is just a degree 5 POLYNOMIAL! $P_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}$. To be fair, the convergence is only fast enough for the first three terms to be sufficient if x is near zero, but if you use $n = 100$ terms, then for all but very large x-values will $\sin x$ be very close to $P_{100}(x) = \sum_{k=0}^{100} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$

Definition The *Taylor series expansion* of $f(x)$ about a point $x = a$ is given by the formula

$$f(x) \ = \ \sum_{k=0}^{\infty} \frac{f^{(k)}(a)(x-a)^k}{k!}.$$

In the above, the notation $f^{(k)}(a)$ means the kth derivative of the function $f(x)$, evaluated at the point $x = a$.

Note: The *MacLaurin series* is simply the Taylor series expanded around $x = 0$. The Taylor polynomial P_n just means truncating the Taylor series at some n-value making the series *finite*, aka a ∗polynomial∗.

Remark The point $x = a$ is called the center of expansion. The center is important because nearby the center you have very fast convergence of the Taylor series to the function it represents. Further away from the center you get slower convergence, or possibly if you get too far away from the center, the series can diverge.

Remark The first degree Taylor polynomial $P_1(x)$ is just $L(x)$ the linearization of the function at $x = a$. The second degree Taylor polynomial is essentially the best quadratic approximation of the function at the point $x = a$. For higher degree Taylor polynomials you get even better approximations, and in the limit, as n tends to infinity, you get the function itself.

Remark I just can't stress how incredibly useful this is. You get to know basically everything there is to know about a function and do all sorts of clever tricks, and all you have to know is all the derivatives of the function *at one single point* $x = a$. It staggers the mind.

Example: Let's approximate $\int_0^1 e^{-x^2}\,dx$ using Taylor polynomials.

Since $e^{-x^2} = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{k!}$, which we can obtain from the MacLaurin series for e^x after replacing $-x^2 \mapsto x$, we can then integrate the series instead of the function

$$\int_0^1 e^{-x^2}\,dx = \int_0^1 \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{x^{2k+1}}{2k+1}|_0^1 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(2k+1)}.$$

We can then approximate the value of $\int_0^1 e^{-x^2}\,dx$ by taking several of the terms of the series. For example, with 5 terms we get

$$\begin{aligned}\sum_{k=0}^{4} \frac{(-1)^k}{k!(2k+1)} &= \frac{1}{0!\cdot 1} - \frac{1}{1!\cdot 3} + \frac{1}{2!\cdot 5} - \frac{1}{3!\cdot 7} \\ &= 1 - \frac{1}{3} + \frac{1}{10} - \frac{1}{42} \\ &= 0.74285714\ldots\end{aligned}$$

For example, using left-endpoint approximation Riemann sum with $n = 200$ rectangles you get 0.7484029014, which is off by only 0.0055457585.

Remark In fact, for alternating sums which satisfy that the terms are decreasing and converging to zero, the next term is an upper bound on the current error.

$$|L - \sum_{k=0}^{N}(-1)^k a_k| < a_{N+1}$$

1.17 Riemann Sum and Taylor Series Problems

Problem: (*Applying the Definite Integral*) A marketing company is trying a new campaign. The campaign lasts for three weeks, and during this time, the company finds that it gains customers as a function of time according to the formula:

$$C(t) = 5t - t^2,$$

where t is time in weeks and the number of customers is given in thousands. Using the general form of the definite integral,

$$\int_a^b f(x)dx = \lim_{n\to\infty} \frac{b-a}{n} \sum_{i=1}^{n} f(x_i^*),$$

calculate the **average** number of customers gained during the three-week campaign.

Solution: First, note that since average value is defined as $AV = \frac{1}{b-a}\int_a^b f(x)dx$, we can use the Riemann sum formula to obtain (the term $b-a$ will cancel):

$$AV = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} C(x_i^*).$$

In this problem, $a = 0$ and $b = 3$. Breaking the interval into n equal pieces would give $\Delta x = \frac{3}{n}$. To find each right-hand endpoint, we can set:

$$x_i^* = a + i\Delta x = 0 + \frac{3i}{n} = \frac{3i}{n},$$

and thus

$$C(x_i^*) = 5\left(\frac{3i}{n}\right) - \left(\frac{3i}{n}\right)^2 = \frac{15}{n}i - \frac{9}{n^2}i^2.$$

Now plugging into the summation:

$$\sum_{i=1}^{n} C(x_i^*) = \sum_{i=1}^{n}\left(\frac{15}{n}i - \frac{9}{n^2}i^2\right) = \frac{15}{n}\sum_{i=1}^{n} i - \frac{9}{n^2}\sum_{i=1}^{n} i^2 = \frac{15}{n}\cdot\frac{n(n+1)}{2} - \frac{9}{n^2}\cdot\frac{n(n+1)(2n+1)}{6}$$

$$= \frac{15(n+1)}{2} - \frac{9(n+1)(2n+1)}{6n}.$$

Using this expression, we can now find average value:

$$AV = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} C(x_i^*) = \lim_{n\to\infty} \frac{1}{n}\left(\frac{15(n+1)}{2} - \frac{9(n+1)(2n+1)}{6n}\right)$$

$$= \lim_{n\to\infty}\left(\frac{15(n+1)}{2n} - \frac{9(n+1)(2n+1)}{6n^2}\right) = \frac{15}{2} - \frac{18}{6} = 4.5,$$

so the company gained an average of 4,500 customers weekly during the campaign.

□

Problem: You are driving when all of a sudden, you see traffic stopped in front of you. You slam the brakes to come to a stop. While your brakes are applied, the velocity of the car is measured, and you obtain the following measurements:

Time since applying breaks (sec)	0	1	2	3	4	5
Velocity of car (in ft/sec)	88	60	40	25	10	0

Using the points given, determine upper and lower bounds for the total distance traveled before the car came to a stop.

Problem: Consider the function $f(x) = x + 2x^2$ on the interval $[0, 2]$. Using a midpoint estimate with $n = 4$ subintervals, estimate the *average value* of f.

Problem: Use a Taylor polynomial to estimate the value of $\sqrt{e}$ with an error of at most 0.01. HINT: Choose $a = 0$ and use the fact that $e < 3$.

Solution: Recall that the remainder term can be found with the formula:

$$|R_n(x)| \le \max|f^{(n+1)}(c)|\frac{|x-a|^{n+1}}{(n+1)!}.$$

Letting $f(x) = e^x$, we have $a = 0$, $x = \frac{1}{2}$, and $f^{(n+1)}(c) = e^c < 3$ for any value of n. So:

$$|R_n(0.5)| \le 3 \cdot \frac{(0.5)^{n+1}}{(n+1)!}.$$

Plug in values of n until this number is smaller than 0.01. Note that $n = 3$ is the first value that works, so we will approximate the value with a 3rd degree Taylor Polynomial. For $f(x) = e^x$, $P_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!}$, so:

$$\sqrt{e} \approx f(0.5) = 1 + 0.5 + \frac{(0.5)^2}{2} + \frac{(0.5)^3}{6} = 1.6458.$$

(the calculator value is approximately 1.64872)

□

Problem: For what values of x can we replace $\cos x$ with $1 - \frac{x^2}{2!} + \frac{x^4}{4!}$ within an error range of no more that 0.001?

Solution: We note that:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ...$$

is an alternating series. For an alternating series, recall that the error is no greater than the absolute value of the **next term** in the sequence.
Thus, we need to solve: $\left|\frac{x^6}{6!}\right| < 0.001$ for x, or $|x^6| < \frac{6!}{1000} = 0.72$, so $|x| < 0.9467$. Choosing x so that $x \in (-0.9467, 0.9467)$ will then give us an approximation within this error range.

□

Problem: Find $f^{(7)}(0)$ for the function $f(x) = x\sin(x^2)$.
Solution: First, find a MacLaurin Series for $f(x) = x\sin(x^2)$:

$$\sin x = \sum(-1)^k \frac{x^{2k+1}}{(2k+1)!} \to \sin(x^2) = \sum(-1)^k \frac{x^{4k+2}}{(2k+1)!},$$

so:

$$x\sin(x^2) = \sum(-1)^k \frac{x^{4k+3}}{(2k+1)!}.$$

Note that in a MacLaurin Series, the coefficient of x^k is exactly $\frac{f^{(k)}(0)}{k!}$, so we need to find the coefficient of x^7. Note that we plug in $k = 1$ to obtain $4k+3 = 7$, so the entire term when $k = 1$ is: $(-1)^1\frac{x^7}{3!} = -\frac{1}{6}x^7$. Then, the coefficient of x^7 is $-\frac{1}{6}$, so:

$$f^{(7)}(0) = 7! \cdot \left(-\frac{1}{6}\right) = -840.$$

□

Chapter 2

Part 2: Linear Algebra (6.5 hours)

Linear Algebra in simplest terms is the study of how to solve systems of linear equations. A linear equation often takes the form $y = mx + b$ where m is the *slope*, b is the *y-intercept*, and the variables x, y are the independent and dependent variables. But in a lot of cases when you are presented with several linear equations that you are supposed to solve simultaneously they are presented in a block format like this.

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= b_2 \\ \vdots \qquad\qquad\qquad\qquad & \quad \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= b_m \end{aligned}$$

The x_i's are the variables, the numbers a_{ij} are called coefficients, and the b_i's are the constants or sometimes they are called the *constraints*.

You can represent the same information as a matrix equation

$$A\mathbf{x} = \mathbf{b},$$

where

$$A = [a_{ij}] = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2,} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Matrix multiplication of AB between an $m \times r$ matrix A, and an $r \times n$ matrix B, is computed as follows: to obtain the entry a_{ij} in the ith row and the jth column of the product AB, you take the ith row of A and the jth column of B

and perform a dot product like operation (it is exactly the dot product if you transpose the row into a column)

$$\begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,r} \end{bmatrix} \begin{bmatrix} b_{1,j} \\ b_{2,j} \\ \vdots \\ b_{r,j} \end{bmatrix} = \sum a_{i,k} b_{k,j}$$

In the sum above, the index of summation is assumed to be the repeated index, but if you like it is

$$[AB]_{i,j} = \sum_{k=1}^{r} a_{i,k} b_{k,j}$$

Note: Matrix multiplication is only defined for matrices of appropriately comparable sizes. In particular, the product AB is defined if the number of columns of A matches the number of rows of B. Hence, AB and BA are *both* defined only if they are square $n \times n$ matrices.

Remark Matrix multiplication is associative and distributive, but not commutative. So $AB \neq BA$ is possible even when both are defined.

Definition The *augmented matrix* of a linear system $A\mathbf{x} = \mathbf{b}$ is the matrix $[A|b]$. Clearly, representing the information as the augmented matrix $[A|b]$, the matrix equation $A\mathbf{x} = \mathbf{b}$, or as a list of linear equalities in block format, are all equivalent ways of storing the data of the linear system of equations.

2.1 Solving Systems of Linear Equations

As mentioned in Part 1, there are many ways to solve systems of linear equations. When there are many equations and many variables, however, it is simplest to use row operations on matrices to simplify the equations before solving.

Definition There are three *elementary row operations* that you can do on a matrix A.

1. Exchange two rows of A,

2. Multiply a row of A by a non-zero constant,

3. Add a multiple of one row of A to another row of A.

It is only slightly non-obvious that all three row operations do not change the solutions of the system of equations represented by the augmented matrix $[A|b]$.

Performing Gaussian-Elimination on a matrix means doing an algorithm by which you systematically reduce a matrix using row operations, until you get a simpler matrix for which it is easy to read off the solutions of the system of linear equations that the matrix represents.

Example: The matrices

$$[A|b] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & \text{-1} \end{array}\right], \text{ and } [B|\hat{b}] = \left[\begin{array}{ccc|c} 7 & 2 & 1 & 24 \\ 3 & 1 & 0 & 11 \\ 6 & 2 & 1 & \text{-21} \end{array}\right]$$

are actually row-equivalent (I made the second from the first in three row operations). So they have the exact same solutions. From looking at the system $A\mathbf{x} = \mathbf{b}$ we see that the solutions are $x_1 = 3, x_2 = 2, x_3 =$ -1.

□

Note: Clearly for the matrix on the left $[A|b]$ it is much easier to determine the solutions. We state the properties that are desirable to be able to read of the solutions below.

Definition A matrix is in *row reduced echelon form* if the following properties are satisfied.

1. All rows of zeros (if any) are at the bottom.

2. The leading entries of each row are 1.

3. The leading ones form a staircase, so that each 1 is above and to the left of the 1's below it.

4. Any entry above a leading 1 is zero.

A matrix which satisfies all but the last property is said to be in *row echelon form*, but it is not reduced.

Example: We will solve the system by row reducing the augmented matrix $[A|b]$

$$\begin{aligned} x_1 + 2x_2 \quad +3x_3 - 2x_4 \quad &= 1 \\ -3x_1 - 6x_2 \quad -9x_3 + 7x_4 \quad &= 0 \\ -2x_1 - 4x_2 \quad -6x_3 + 5x_4 \quad &= 1 \end{aligned}$$

Doing the row operations

$$\begin{aligned} 3R_1 + R_2 &\mapsto R_2 \\ 2R_2 + R_3 &\mapsto R_3 \\ -R_2 + R_3 &\mapsto R_3 \\ 2R_2 + R_1 &\mapsto R_1 \end{aligned}$$

yields the rref of $[A|b]$ which is

$$[A|b] \sim \left[\begin{array}{cccc|c} 1 & 2 & 3 & 0 & 7 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Thus the solutions of $[A|b]$ are the same as the solutions of

$$\begin{aligned} x_1 + 2x_2 \quad +3x_3 \quad &= 1 \\ x_4 \quad &= 3 \\ 0 \quad &= 0 \end{aligned}$$

By convention we write the equations in terms of the non-pivot variables, and give any non-pivot variables new parameter names. So the solutions are given by

$$\begin{aligned} x_1 &= -2r - 3r + 1 \\ x_2 &= r \quad \text{(free)} \\ x_3 &= s \quad \text{(free)} \\ x_4 &= 3 \end{aligned}$$

□

2.2 Determinants

Definition To compute the *determinant* of an $n \times n$ matrix you perform cofactor expansion across a row or column. There are several equivalent notations which specify how to do this.

$$\det(A) = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} A_{1n} = \sum_{j=1}^{n} (-1)^{1+j} \det A_{1j}$$

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots a_{nj}C_{nj} = \sum_{i=1}^{n} a_{ij}C_{ij}$$

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}$$

In the above, A_{ij} means the matrix obtained from A by deleting row i and column j, and is called the ijth-minor of A. The notation C_{ij} is called the *cofactors* of A and is defined by $C_{ij} = (-1)^{i+j} \det A_{ij}$.

Example: Compute the determinant of A using a few columns or rows.

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix} \quad \square$$

2.3 Workshop 3: Matrix algebra and linear system solving

1. In the questions below, assume that A is a (general) $m \times n$ matrix, with m rows and n columns, and b is a (general) vector in $\mathbb{R}^m$. Circle TRUE if the statement is **always** true (for any possible choice of A and b), otherwise circle FALSE. (4 pts. each)

(a) If $m > n$, then $Ax = b$ has a unique solution. TRUE FALSE

(b) If $m = n$, then $Ax = b$ has a unique solution. TRUE FALSE

(c) If $Ax = b$ has a unique solution, then there are n pivots in A. TRUE FALSE

(d) If there are n pivots in A, then $Ax = b$ has a unique solution. TRUE FALSE

(e) If $Ax = 0$ has more than one solution, then it has infinitely many solutions. TRUE FALSE

(f) The matrix $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ is in rref. TRUE FALSE

(g) The matrix $\begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is in rref. TRUE FALSE

2. Find the rref (reduced row echelon form) of the matrix

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ -2 & 5 & -12 & 11 \\ 1 & -2 & 5 & -4 \end{bmatrix}$$

3. Determine whether or not the vector equation below has a solution. If it has a unique solution, find it. If it has infinitely many solutions express the solutions parametrically in terms of the free variable(s).

$$x \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} + z \begin{bmatrix} 2 \\ 11 \\ 8 \end{bmatrix} = \begin{bmatrix} -1 \\ 8 \\ 5 \end{bmatrix}$$

4. Suppose v_1, v_2 are two vectors in $\mathbb{R}^2$, and b is another vector in $\mathbb{R}^2$. Give an example of vectors v_1, v_2, and b such that the vector equation $xv_1 + yv_2 = b$ has

(a) A unique solution.

(b) No solution.

(c) Infinitely many solutions.

5. Find the determinant of A.

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

6. Solve.

$$\begin{aligned} 2x - 3y &= 7 \\ x + 5y &= 3 \end{aligned}$$

7. Solve.

$$\begin{aligned} x - 2y + z &= 1 \\ 2y - z &= 3 \\ x \qquad - z &= 2 \end{aligned}$$

8. True or false. Assume the matrix A has 3 rows and 4 columns, so it's size is 3×4, meaning the corresponding system has 3 equations and 4 unknowns.

(a) TRUE/FALSE If A has three pivot positions, then the equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution.

(b) TRUE/FALSE If A has three pivot positions, then the equation $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.

(c) TRUE/FALSE If $\mathbf{x}$ is a nontrivial solution to $A\mathbf{x} = \mathbf{0}$, then every entry in $\mathbf{x}$ is nonzero.

(d) TRUE/FALSE If A has three pivot positions, then the equation $A\mathbf{x} = \mathbf{b}$ is always consistent for all $\mathbf{b} \in \mathbb{R}^3$.

(e) TRUE/FALSE The homogeneous system $A\mathbf{x} = 0$ has infinitely many solutions.

9. In $\mathbb{R}^3$ (so using three coordinate axes) sketch the following, making a new sketch for each part (i)-(v):

(i) the plane $z = 0$,

(ii) the plane $z = 2$,

(iii) the plane $y = -3$,

(iv) the plane $x + y + z = 0$,

(v) the intersection of the planes $z = 2$ and $x = 0$.

In each case, you are practicing drawing an accurate, representative graph of the plane of points which satisfy the given equation in the variables x, y, and z.

10. (1) Choose two vectors v_1, v_2 in $\mathbb{R}^2$ and another vector b also in $\mathbb{R}^2$. Find scalars x, y in $\mathbb{R}$ such that $xv_1 + yv_2 = b$ (if this is not possible, pick other v_1, v_2, b vectors). Illustrate the vector equation you just solved by graphing the vectors v_1, v_2, b in the $x - y$-plane, and be sure to illustrate how b is obtained by adding a scalar of one vector to the other. (2) Repeat part 1 with vectors v_1, v_2, b in $\mathbb{R}^3$ that give a consistent system. (3) Why is part 2 more difficult than part 1? Explain clearly using complete sentences.

2.4 Matrix Algebra

Once you've defined matrix multiplication and matrix addition (which is just adding componentwise), it makes sense to look for additive and multiplicative identity objects in the set of square matrices. In the real numbers, the additive identity is 0, and the multiplicative identity is 1. That's because

1. $0 + a = a$ for any $a \in \mathbb{R}$, and

2. $1 \cdot a = a$ for any $a \in \mathbb{R}$.

The corresponding matrices that do the same "jobs" as 0 and 1 for the real numbers are the zero matrix (which is just denoted 0) and the identity matrix which is denoted I.

The identity matrix takes the form

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

for the 2×2 identity matrix. And

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

for the 3×3 identity matrix. In general, for any n, the matrix I_n with 1's along the diagonal and 0's elsewhere is the $n \times n$ identity matrix.

Every matrix A has an additive inverse $-A$, which is just the matrix which gets you to the additive identity 0. So $A + (-A) = A - A = 0$.

However, NOT every matrix has a *multiplicative* inverse. For a given matrix A, sometimes it is possible to find a pair of matrices $AB = I$, but sometimes $AB = I$ is impossible.

Definition Given an $n \times n$ matrix A, we say that A is *invertible* if $AB = I$ for some matrix B. In that case we call B the inverse of A and write $B = A^{-1}$.

Note: We have the following properties if A is an $n \times n$ invertible matrix with inverse matrix A^{-1}.

1. $(A^{-1})^{-1} = A$.

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

3. $I^{-1} = I$.

4. $A^{-1}A = AA^{-1} = I$

5. If $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x} = A^{-1}\mathbf{b}$.

Note: The last item on the previous list gives you an easy way to compute the solution to a system of linear equations if A is invertible.

Theorem The inverse of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by the formula

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

Theorem If A is invertible, then the rref of A is the identity matrix I.

The way to find inverses is to row reduce the matrix A to I but keeping track of the row operations you are using to reduce the matrix A. If you perform the *exact same* operations on I then you will get A^{-1}.

Remark To find A^{-1}.

Step 1: Form the augmented matrix $[A|I]$.

Step 2: Perform Gaussian elimination to row reduce the augmented matrix to obtain $[I|A^{-1}]$.

Step 3: Read off the inverse matrix as the second part of the augmented matrix.

Example: The matrix

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 0 \\ 2 & 6 & 5 \end{bmatrix}$$

has inverse

$$A^{-1} = \begin{bmatrix} 5 & \text{-}3 & \text{-}2 \\ 0 & 1 & 0 \\ \text{-}2 & 0 & 1 \end{bmatrix}.$$

□

2.5 Vector Spaces

You've actually been working with vector spaces since grade school, although they are rarely called that. The xy-plane is a vector space, in the sense that every point in the xy-plane represents a vector (and visa versa). Also, the real line is the simplest example of a vector space (somewhat trivially). Polynomials of degree at most d form a vector space because of the following two properties: (1) if you take two polynomials of degree at most d and add them together, you get a polynomial of degree at most d, and (2) if you multiply a polynomial of degree at most d by a scalar, you get another polynomial of degree at most d.

Definition A *vector space* is a set of vectors V which satisfy the properties of being closed under vector addition and closed under scalar multiplication.

1. If $\mathbf{x}, \mathbf{y} \in V$ then $\mathbf{x} + \mathbf{y} \in V$.

2. If $\mathbf{x} \in V$ and $c \in \mathbb{R}$, then $c\mathbf{x} \in V$.

Definition A *subspace* W of a vector space V is a subset $W \subseteq V$ of V that is also itself a vector space.

Example: Any line passing through the origin is a vector subspace of $\mathbb{R}^2$. However, the first quadrant of $\mathbb{R}^2$ is not a subspace, it is not closed under scalar multiplication. The union of the first and third quadrant is not a subspace either, because although it is closed under scalar multiplication, it is not closed under vector addition.

□

Example: Polynomials of degree at most d are a vector space for any $d \geq 0$. Polynomials of degree *exactly* d are *not* vector spaces for any d except $d = 0$. (WHY NOT?)

□

Example: The solutions $\mathbf{x}$ to a homogeneous system of linear equations $A\mathbf{x} = 0$ form a vector subspace, since $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x}+A\mathbf{y} = 0+0 = 0$ (provided $\mathbf{x}$ and $\mathbf{y}$ are both solutions to the homogeneous system), and $A(c\mathbf{x}) = c(A\mathbf{x}) = c(0) = 0$ (if $\mathbf{x}$ is a solution to the homogeneous system). (hence the solutions are closed under vector addition and scalar multiplication).

□

Definition The vector subspace which is the set of solutions of a homogeneous system of linear equations defined by $A\mathbf{x} = 0$ is denoted

$$\text{nul}(A) = \{\mathbf{x} \mid A\mathbf{x} = 0\}$$

and is called the null space of A.

Definition The span of the columns of a matrix A is a subspace and is called the *column space* of A, and denoted $\text{col}(A)$.

In vector spaces, two of the important concepts are *span* and *linear independence*. In words, the span of a set of vectors are all the other vectors you can get from adding up scalar multiples of the vectors. So, the span of a single vector is the line containing that vector (which necessarily passes through the origin). A set of vectors is linear independent if it is a minimal spanning set of vectors, meaning that if you deleted any vectors from the spanning set you would make the span smaller. We make these notions precise below.

Definition A set of vectors $B = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^m$ is *linearly independent* if whenever

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0,$$

it must be that

$$c_1 = c_2 = \cdots = c_n = 0.$$

Note: This is equivalent to saying that the $m \times n$ matrix

$$A = [v_1 \ v_2 \ \cdots \ v_n]$$

has the property that $A\mathbf{x} = 0$ has only the trivial solution $\mathbf{x} = 0$.

Definition Given a set of $B = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^m$, we say that B spans W if every vector in W can be written as a linear combination of the vectors in B. In symbols, for every $\mathbf{w} \in W$, there exist $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{w}$$

Note: This is equivalent to saying that the matrix equation $A\mathbf{x} = \mathbf{w}$ has a solution $\mathbf{x}$ for every $\mathbf{w} \in W$.

Remark It is not too difficult to show that: Any two linearly independent vectors in $\mathbb{R}^2$ span all of $\mathbb{R}^2$; Any three vectors in $\mathbb{R}^2$ are automatically linearly dependent; Any set of vectors which includes the zero vector is a linearly dependent set of vectors.

Definition If a set of vectors B is linearly independent and spans a subspace W of a vector space V, then we say that B is a basis for W.

Remark The standard basis for $\mathbb{R}^2$ is $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. The standard basis for $\mathbb{R}^n$ is $e_1, e_2, \ldots, e_n$ where e_i has an 1 in the ith row and 0's elsewhere.

Theorem Given a square matrix A which is $n \times n$. The following are equivalent statements:

1. The columns of A are linearly independent.

2. The columns of A span $\mathbb{R}^n$.

3. The determinant of A is non-zero.

4. The rref of A is the identity matrix I which has 1's along the diagonal and 0's elsewhere.

2.6 Workshop 4: Matrix algebra and vector spaces

1. Express the vector $\begin{bmatrix} -5 \\ -2 \\ -1 \end{bmatrix}$ as a linear combination of the vectors

$$\begin{bmatrix} 9 \\ 0 \\ 9 \end{bmatrix}, \quad \begin{bmatrix} 6 \\ 3 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}.$$

2. Solve the system below by writing $A\mathrm{x} = b$, find A^{-1} and use it to solve the system for x.

$$\begin{aligned} -x_1 + 3x_2 \quad &= -1 \\ -x_1 - 6x_2 \quad &= \ \ 2 \\ 2x_2 \ \ -x_3 &= -1 \end{aligned}$$

3. Are the vectors linearly independent?

$$\left\{ \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

4. In the questions below, assume that A is a (general) $m \times n$ matrix, with m rows and n columns, and b is a (general) vector in $\mathbb{R}^m$. Circle TRUE if the statement is **always** true (for any possible choice of A and b), otherwise circle FALSE.

5. If $Ax = b$ has a unique solution, then b is in the span of the columns of A. TRUE FALSE

6. If $m > n$, then the columns of A are linearly independent. TRUE FALSE

7. If $m < n$, then the columns of A are linearly dependent. TRUE FALSE

8. Find the value for h that makes the given vectors linearly dependent.

$$\begin{bmatrix} 4 \\ -2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ h \end{bmatrix}.$$

9. Determine whether the given set of vectors is linearly independent/dependent. No justification is necessary for full credit.

(a) $\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$

(b) $\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} \right\}$

(c) $\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} \right\}$

10. Find the solutions of the matrix equation $A\mathbf{x} = 0$ where A is the matrix below. Please write your answer in parametric **vector** form.

$$A = \begin{bmatrix} 1 & 0 & -3 & 2 & 0 \\ -1 & 0 & 3 & -2 & 1 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note: In this problem, parametric vector form means to express the solutions to this problem as a linear combination of vectors with the free variables as the scalars. This is meant to show you by example that the solutions to $Ax = 0$ are a subspace of $\mathbb{R}^n$, since they are a span.

11. Determine whether the given vectors are linearly independent or linearly dependent. If the vectors are linearly dependent find a non-trivial linear combination of the vectors which give the zero vector.

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \\ 2 \end{bmatrix}$$

12. For each part, if possible, give an example of two sets A and B of vectors $\{v_1, v_2, \ldots, v_n\}$ in $\mathbb{R}^m$ where the set A is **linearly independent** and the set B is **linearly dependent**, and if it is not possible to do so for either A or B explain why in your own words.

1. One vector in $\mathbb{R}^2$,

2. two vectors in $\mathbb{R}^2$,

3. three vectors in $\mathbb{R}^2$,

4. two vectors in $\mathbb{R}^3$,

5. three vectors in $\mathbb{R}^3$,

6. four vectors in $\mathbb{R}^3$.

7. For each problem find matrices which satisfy the given conditions. You don't have to justify *how* you found the matrices for each problem, but you *must verify the equality with calculations* in each case.

(a) $AB = BA$ but neither A nor B is 0 nor I.

(b) $AB \neq BA$.

(c) $AB = AC$ but $B \neq C$.

(d) $AB = 0$ but neither A nor B is 0.

(e) $AB = I$ but neither A nor B is I.

2.7 Linear Transformations

A linear transformation is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ whose inputs and outputs are vectors, which satisfies the linearity conditions

1. $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$, and

2. $T(c\mathbf{x}) = cT(\mathbf{x})$.

Note: Notice that defining $T(\mathbf{x}) = A\mathbf{x}$ for an appropriately sized matrix A defines a linear transformation. It turns out that these are the *only* linear transformations as made precise in the following definition.

Definition Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we define the *standard matrix* of T to be $A = [T(e_1)\ T(e_2)\ \ldots\ T(e_n)]$. Notice that the size of A is $m \times n$.

The notions of span and linear independence have corresponding meanings for linear transformations. In particular

Note: If the columns of A are linearly independent, then the linear transformation defined by $T_A(\mathbf{x}) = A\mathbf{x}$ is one-to-one, meaning that $T(\mathbf{x}) = T(\mathbf{y})$ implies $\mathbf{x} = \mathbf{y}$. Furthermore, if the columns of A span a vector subspace W, then we say that T is onto W.

Definition A linear transformation $T : V \to W$ is said to be an *isomorphism* if it is a one-to-one map and it is onto W.

Note: The word “map” is synonymous with “transformation” or “function”.

2.8 Linear Transformation Problems

1. For each matrix A below, (0) state the domain and codomain of T_A, (1) find $T_A(e_1), T_A(e_2)$, (2) find $T_A(v), T_A(w)$, (3) describe in a few words what the transformation is doing, and (4) give the matrix an appropriate "name". For the problems below use

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad w = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

1. $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

2. $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

3. $A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

5. $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

7. $A = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$

Now, for the problems below use

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad w = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$$

1. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

2. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

3. Consider the linear transformation T which first rotates vectors in $\mathbb{R}^2$ by 90° counterclockwise, then reflects the result about the x-axis. Find the standard matrix A of T as well as the image $T\left(\begin{bmatrix}1\\1\end{bmatrix}\right)$. *Hint: the first column of* A *is* $T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)$ *and the second column of* A *is* $T\left(\begin{bmatrix}0\\1\end{bmatrix}\right)$.

4. Is the transformation $T(\mathbf{v}) = \mathbf{v} + \begin{bmatrix}1\\1\end{bmatrix}$ a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$? Explain.

5. For each of the above problems (where a matrix A was defined), find a basis for the the null space $\text{nul}(A)$ and the column space $\text{col}(A)$.

2.9 Eigenvectors and Eigenvalues

One of the most important concepts in linear algebra is the notion of eigenvectors and eigenvalues. The eigenvalues of a matrix tell you a lot of information about the matrix, and you can actually fully recover an $n \times n$ matrix A if you know all its eigenvalues and there is a basis of $\mathbb{R}^n$ consisting of its eigenvectors (this is called *diagonalization*).

Definition Given a square $n \times n$ matrix A, we say that $\lambda \in \mathbb{R}$ is an eigenvalue of A with associated eigenvector $\mathbf{v} \in \mathbb{R}^n$ if $\mathbf{v} \neq 0$ and

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Note: The condition that at least the eigenvector $\mathbf{v}$ is non-zero is important because if we allow the zero vector 0 to be an eigenvector then *every number λ would be an eigenvalue associated to this eigenvector.* But we want eigenvalues to be special, so that's why we only allow non-zero vectors to be associated eigenvectors.

Note: However, we want to use the fact that the set of eigenvectors associated to a particular eigenvalue is a subspace, so we define the *eigenspace associated to an eigenvector* λ of a matrix A to be the set of all eigenvectors of A associated to λ *and* the zero vector (because any vector space must include 0).

Remark Note that if $A\mathbf{v} = \lambda\mathbf{v}$ and $\mathbf{v} \neq 0$, then $(A - \lambda I)\mathbf{x} = 0$ has a non-zero solution, hence the matrix $A - \lambda I$ has determinant equal to zero (because if the determinant was non-zero then this matrix would be invertible, and the only solution to the homogeneous equation would be the trivial solution). In fact, this is exactly how we find eigenvalues and eigenvectors.

Theorem The eigenvalues of an $n \times n$ matrix A are the solutions λ to setting the *characteristic polynomial* to zero

$$p(\lambda) = \det(A - \lambda I) = 0.$$

The eigenvectors of A are the vectors in the null space of $A - \lambda I$, *i.e.*, the solution vectors to

$$(A - \lambda I)\mathbf{x} = 0.$$

2.10 Diagonalization

If an $n \times n$ matrix A has a basis of $\mathbb{R}^n$ which consists entirely of eigenvectors, then A is *diagonalizable.* This means that

$$P^{-1}AP = D$$

for a diagonal matrix D and an invertible matrix P. It turns out that

$$P = [v_1\ v_2\ \dots\ v_n], \quad D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Note: The operation $P^{-1}AP$ is called a change of basis because that's what multiplying by P, then A, then P^{-1} does. For example, let's see what $P^{-1}AP$ does to e_1.

$$P^{-1}APe_1 = P^{-1}Av_1 = P^{-1}\lambda_1 v_1 = \lambda_1 e_1.$$

So this transformation *makes the standard basis vectors into eigenvectors.* Another way to think about it is that using the eigenvectors $\{v_1, \dots, v_n\}$ as your basis for $\mathbb{R}^n$ (instead of the standard basis vectors $e_1, \dots, e_n$) makes your matrix A act like a diagonal matrix (which just scales the coordinate axes).

Note: It is sometimes useful to compute a diagonalization of a matrix A in order to easily compute a large power of A. For instance, if A is diagonalizable then

$$A^{100} = (PDP^{-1})^{100} = PD^{100}P^{-1}.$$

You should make sure you understand which of the above equalities is non-trivial. Also, make sure you understand why A^{100} is more difficult to compute than D^{100} if D is diagonal (and A is not).

2.11 Workshop 5: Eigenvectors and eigenvalues

1. The vector $v = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}$ is an eigenvector of the matrix

$$A = \begin{bmatrix} 2 & 0 & -4 & 0 \\ -2 & 3 & -8 & 0 \\ 0 & 0 & -6 & 0 \\ 0 & 0 & -3 & 0 \end{bmatrix}.$$

Find the eigenvalue λ associated to v. *Hint: write down the definition of eigenvalue.*

2. If A is the 2×2 matrix with eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 3$ and corresponding eigenvectors

$$v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix},$$

then what is A? *Hint: Diagonalize!*

3. Each of the following questions is answered with a number(s), matrix, or vector.

(a) Find all eigenvalues of $\begin{bmatrix} 2 & 0 & 3 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

(b) Find an eigenvector of $\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}$ corresponding to the eigenvalue $\lambda = 2$.

(c) List an eigenvector of the standard matrix of the linear transformation which reflects the standard basis vectors in $\mathbb{R}^2$ across the line $y = -x$.

(d) If the eigenvalues of A are $1, 2$, and 3, then what are the eigenvalues of A^3?

4. In the questions below, assume that A is a square $n \times n$ matrix. Circle TRUE if the statement is **always** true, otherwise circle FALSE.

(a) If A has n linearly independent eigenvectors, then A has n distinct eigenvalues. TRUE FALSE

(b) The first standard basis vector e_1 is an eigenvector TRUE FALSE
of the standard matrix of the linear transformation which rotates $\mathbb{R}^2$ by 180°.

(c) If A is diagonalizable, then A is invertible. TRUE FALSE

(d) The vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector TRUE FALSE
of the matrix $\begin{bmatrix} 2 & 0 \\ -2 & 3 \end{bmatrix}$.

(e) If zero is an eigenvalue of A, then A is not invertible. TRUE FALSE

(f) If A is not invertible, then zero is an eigenvalue of A. TRUE FALSE

(g) If $A = PBP^{-1}$, then $\det(A) = \det(B)$. TRUE FALSE

(h) If λ is an eigenvalue of A, and B is the reduced row echelon form of A, then λ is an eigenvalue of B too. TRUE FALSE

Chapter 3

Part 3: Probability and Statistics (4 hours)

Will assume that you know about sample spaces, events, and the definition of probability.

3.1 Basics: Conditional Probability

Definition: $P(A|B) \equiv P(A \cap B)/P(B)$ is the *conditional probability of A given B.*
Example: Toss a fair die. Let $A = \{1, 2, 3\}$ and $B = \{3, 4, 5, 6\}$. Then

$$P(A|B) \;=\; \frac{P(A \cap B)}{P(B)} \;=\; \frac{1/6}{4/6} \;=\; 1/4. \quad \Box$$

The idea of conditional probability is that you are sampling out of a specific set of events, so your sample space is smaller. If the knowledge of whether B happened does not effect the outcome of A, then we say that A and B are independent events.
Definition: If $P(A \cap B) = P(A)P(B)$, then A and B are *independent* events.
It follows from the definitions that
Theorem: If A and B are independent, then $P(A|B) = P(A)$.
It can be easily computed whether two events are independent or not using this theorem.
Example: Toss two dice. Let $A =$ "Sum is 7" and $B =$ "First die is 4". Then

$$P(A) = 1/6, \quad P(B) = 1/6, \quad \text{and}$$

$$P(A \cap B) = P((4,3)) = 1/36 = P(A)P(B).$$

So A and B are independent. $\Box$

3.2 Basics: Random Variables

The term random variable is just a formal way of defining the outcome of an experiment. In some texts we allow random variables to take values that are "words" or "colors", or anything else we choose. So in that case we could say X is the random variable of the color of socks in my sock drawer.

For our purposes of this bootcamp, we will only allow random variables to take numeric values. So for instance we could say $X = 1$ if the sock is white and $X = 2$ if the sock is red, etc.

Definition: A *random variable* (RV) X is a function from the sample space Ω to the real line, i.e., $X : \Omega \to \mathbb{R}$.

Example: Let X be the sum of two dice rolls. Then $X((4, 6)) = 10$. In addition,

$$P(X = x) = \begin{cases} 1/36 & \text{if } x = 2 \\ 2/36 & \text{if } x = 3 \\ \vdots & \\ 1/36 & \text{if } x = 12 \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

We assign probabilities to events using a pmf. The pfm needs to satisfy that the values it takes is non-negative and that the sum of all it's values is 1 (where you replace "sum" with "integral" if the random variable is a continuous random variable).

Definition: If the set of possible values of a RV X is finite or countably infinite, then X is a *discrete* RV. Its *probability mass function* (pmf) is $f(x) := P(X = x)$. Note that $\sum_x f(x) = 1$.

Example: Flip 2 coins. Let X be the number of heads.

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

3.3 Well Known Discrete Distributions

Here are some well-known discrete RV's that you may know (from Wikipedia):

1. Bernoulli(p). The discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1?p$; that is, the probability distribution of any single experiment that asks a yes?no question, with pmf

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0 \end{cases}$$

2. Binomial(n, p). The binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking a yes?no question, with pmf

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N.

3. Geometric(p). Based on convention, can be either of the two defintions.

The probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set $\{1, 2, 3, \ldots\}$ The probability distribution of the number $Y = X?1$ of failures before the first success, supported on the set $\{0, 1, 2, 3, \ldots\}$.

These two different geometric distributions should not be confused with each other. Often, the name shifted geometric distribution is adopted for the former one (distribution of the number X); however, to avoid ambiguity, it is considered wise to indicate which is intended, by mentioning the support explicitly.

The geometric distribution gives the probability that the first occurrence of success requires k independent trials, each with success probability p. If the probability of success on each trial is p, then the probability that the kth trial (out of k trials) is the first success is $\Pr(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, 3, \ldots$.

The above form of the geometric distribution is used for modeling the number of trials up to and including the first success. By contrast, the following form of the geometric distribution is used for modeling the number of failures until the first success:

$P(Y = k) = (1-p)^k p$ for $k = 0, 1, 2, 3, \ldots$.

In either case, the sequence of probabilities is a geometric sequence.

Example: For example, suppose an ordinary die is thrown repeatedly until the first time a "1" appears. The probability distribution of the number of times it is thrown is supported on the infinite set $\{1, 2, 3, \ldots\}$ and is a geometric distribution with $p = 1/6$.

4. Negative Binomial.

The negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted r) occurs. For example, if we define a 1 as failure, all non-1s as successes, and we throw a dice repeatedly until 1 appears the third time (r = three failures), then the probability distribution of the number of non-1s that appeared will be a negative binomial distribution.

The probability mass function of the negative binomial distribution is

$f(k; r, p) := P(X = k) = \binom{k+r-1}{k} p^k (1-p)^r \quad \text{for } k = 0, 1, 2, \ldots$

5. Poisson(λ).

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

For instance, an individual keeping track of the amount of mail they receive each day may notice that they receive an average number of 4 letters per day. If receiving any particular piece of mail does not affect the arrival times of future pieces of mail, i.e., if pieces of mail from a wide range of sources arrive independently of one another, then a reasonable assumption is that the number of pieces of mail received in a day obeys a Poisson distribution. Other examples that may follow a Poisson include the number of phone calls received by a call center per hour and the number of decay events per second from a radioactive source.

An event can occur $0, 1, 2, \ldots$ times in an interval. The average number of events in an interval is designated λ (lambda). Lambda is the event rate, also called the rate parameter. The probability of observing k events in an interval is given by the equation

$$P(k \text{ events in interval}) = e^{-\lambda} \frac{\lambda^k}{k!}$$

where λ is the average number of events per interval e is the number $2.71828\ldots$ (Euler's number) the base of the natural logarithms k takes values $0, 1, 2, \ldots$ $k! = k(k-1)(k-2)\ldots(2)(1)$ is the factorial of k. This equation is the probability mass function for a Poisson distribution.

Notice that this equation can be adapted if, instead of the average number of events λ, we are given a time rate r for the events to happen. Then $\lambda = rt$

(with r in units of 1/time), and

$$P(k \text{ events in interval } t) = e^{-rt}\frac{(rt)^k}{k!}$$

Example: On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of k = 0, 1, 2, 3, 4, 5, or 6 overflow floods in a 100-year interval, assuming the Poisson model is appropriate.

Because the average event rate is one overflow flood per 100 years, ? = 1

$$P(k \text{ overflow floods in 100 years}) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1^k e^{-1}}{k!}$$
$$P(k = 0 \text{ overflow floods in 100 years}) = \frac{1^0 e^{-1}}{0!} = \frac{e^{-1}}{1} \approx 0.368$$
$$P(k = 1 \text{ overflow flood in 100 years}) = \frac{1^1 e^{-1}}{1!} = \frac{e^{-1}}{1} \approx 0.368$$
$$P(k = 2 \text{ overflow floods in 100 years}) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2} \approx 0.184$$

The table below gives the probability for 0 to 6 overflow floods in a 100-year period.

k	P(k overflow floods in 100 years)
0	0.368
1	0.368
2	0.184
3	0.061
4	0.015
5	0.003
6	0.0005

□

Example: Ugarte and colleagues report that the average number of goals in a World Cup soccer match is approximately 2.5 and the Poisson model is appropriate.

Because the average event rate is 2.5 goals per match, ? = 2.5.

$$P(k \text{ goals in a match}) = \frac{2.5^k e^{-2.5}}{k!}$$

$$P(k = 0 \text{ goals in a match}) = \frac{2.5^0 e^{-2.5}}{0!} = \frac{e^{-2.5}}{1} \approx 0.082$$

$$P(k = 1 \text{ goal in a match}) = \frac{2.5^1 e^{-2.5}}{1!} = \frac{2.5e^{-2.5}}{1} \approx 0.205$$

$$P(k = 2 \text{ goals in a match}) = \frac{2.5^2 e^{-2.5}}{2!} = \frac{6.25e^{-2.5}}{2} \approx 0.257$$

The table below gives the probability for 0 to 7 goals in a match.

k	P(k goals in a World Cup soccer match)
0	0.082
1	0.205
2	0.257
3	0.213
4	0.133
5	0.067
6	0.028
7	0.010

□

3.4 Continuous Random Variables and cdf's

Continuous distributions are concerned with random variables that take continuous rather than discrete values. For example, temperature, the amount of time for something to happen, or the length of a caterpillar are all continuous random variables.

Definition: A *continuous* RV is one with probability zero at every individual point, and for which there exists a *probability density function* (pdf) $f(x)$ such that $P(X \in A) = \int_A f(x)\,dx$ for every set A. Note that $\int_{\mathbb{R}} f(x)\,dx = 1$.

Here is a contrived example

Example: Pick a random number between 3 and 7. Then

$$f(x) \;=\; \begin{cases} 1/4 & \text{if } 3 \le x \le 7 \\ 0 & \text{otherwise} \end{cases} \qquad \Box$$

Note: Notice that in order to verify the property that the total probability is 1 you can integrate or just draw a picture and use simple geometry.

Notation: Often you will see the notation $X \sim N(0,1)$ which means X is normally distrubuted. In general "$\sim$" means "is distributed as". For instance, $X \sim \text{Unif}(0,1)$ means that X has the uniform distribution on $[0,1]$.

The cumulative distribution function tells you the probability that a random variable's value is *at most* some upper limit. This is a surprisingly useful concept.

Definition: For any RV X (discrete or continuous), the *cumulative distribution function* (cdf) is

$$F(x) \;\equiv\; P(X \le x) \;=\; \begin{cases} \sum_{y \le x} f(y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} f(y)\,dy & \text{if } X \text{ is continuous} \end{cases}$$

Note: Note that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. In addition, if X is continuous, then by the FTC we have $\frac{d}{dx}F(x) = f(x)$.

Here are a couple examples to practice on.

Example: Flip 2 coins. Let X be the number of heads.

$$F(x) \;=\; \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \le x < 1 \\ 3/4 & \text{if } 1 \le x < 2 \\ 1 & \text{if } x \ge 2 \end{cases} \qquad \Box$$

Example: if $X \sim \text{Exp}(\lambda)$ (i.e., X is exponential with parameter λ), then $f(x) = \lambda e^{-\lambda x}$ and $F(x) = 1 - e^{-\lambda x}$, $x \ge 0$. $\Box$

3.5 Well Known Continuous Distributions

Here are some well-known continuous (from Wikipedia) RV's:

1. Uniform(a, b). the continuous uniform distribution or rectangular distribution is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the distribution's support are equally probable.

The pdf of the continuous uniform distribution is

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

The cumulative distribution function is:

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x > b \end{cases}$$

Its inverse is:

$$F^{-1}(p) = a + p(b-a) \quad \text{for } 0 < p < 1.$$

2. Exponential(λ).

the exponential distribution (also known as negative exponential distribution) is the probability distribution that describes the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate. It is a particular case of the gamma distribution. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless. In addition to being used for the analysis of Poisson point processes it is found in various other contexts.

The probability density function (pdf) of an exponential distribution is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \\ 0 & x < 0. \end{cases}$$

Alternatively, this can be defined using the right-continuous Heaviside step function, $H(x) = 1$ if $x \ge 0$ and $H(x) = 0$ if $x < 0$,

Example: if $X \sim \text{Exp}(\lambda)$ (i.e., X is exponential with parameter λ), then $f(x) = \lambda e^{-\lambda x}$ and $F(x) = 1 - e^{-\lambda x}$, $x \ge 0$. □

$$f(x; \lambda) = \lambda e^{-\lambda x} H(x).$$

Here $\lambda > 0$ is the parameter of the distribution, often called the rate parameter. The distribution is supported on the interval $[0, \infty)$. If a random variable X has this distribution, we write $X \sim \text{Exp}(\lambda)$.

3. Normal(μ, σ^2).

The normal distribution is sometimes informally called the bell curve. Formally, a random variable with a Gaussian distribution is said to be normally distributed and is called a normal deviate.

The probability density of the normal distribution is

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

where

- μ is the mean or expectation of the distribution (and also its median and mode),
- σ is the standard deviation, and
- σ^2 is the variance.

Note: The normal distribution is useful because of the central limit theorem. In its most general form, under some conditions (which include finite variance), it states that averages of samples of observations of random variables independently drawn from independent distributions converge in distribution to the normal, that is, become normally distributed when the number of observations is sufficiently large. Physical quantities that are expected to be the sum of many independent processes (such as measurement errors) often have distributions that are nearly normal. Moreover, many results and methods (such as propagation of uncertainty and least squares parameter fitting) can be derived analytically in explicit form when the relevant variables are normally distributed.

3.6 Simulating Simple Random Variables

We'll make a brief aside here to show how to simulate some very simple random variables. So the idea is like this: suppose you are writing code for a game and you need to simulate a dice toss. How do you do it? Or suppose you have some discrete random variable that you want to simulate in a game, so like my socks come out of a video-game-dresser-drawer with probability such and such for each color, how do you program that? Well here is some examples for how to simulate some simple discrete random variables.

Example (Discrete Uniform): Consider the discrete uniform distribution on $\{1, 2, \ldots, n\}$, which is to say $X = i$ with probability $1/n$ for $i = 1, 2, \ldots, n$. (If you are an *old school* gamer this is an n-sided dice toss.)

If $U \sim \text{Unif}(0, 1)$ is a continuous random variable with the uniform distribution on the interval $[0, 1]$, then we can easily create a discrete uniform random variate from U simply by setting $X = \lceil nU \rceil$, where $\lceil\cdot\rceil$ is the "ceiling" (or "round up") function.

For example, if $n = 10$ and we sample a Unif(0,1) random variable $U = 0.73$, then $X = \lceil 7.3 \rceil = 8$.

□

Example: Another example. Consider

$$P(X = x) = \begin{cases} 0.25 & \text{if } x = -2 \\ 0.10 & \text{if } x = 3 \\ 0.65 & \text{if } x = 4.2 \\ 0 & \text{otherwise} \end{cases}$$

This random variable isn't simulated by a die toss. Instead, we use what's called the *inverse transform method*.

x	$f(x)$	$P(X \le x)$	Unif(0,1)'s
-2	0.25	0.25	$[0.00, 0.25]$
3	0.10	0.35	$(0.25, 0.35]$
4.2	0.65	1.00	$(0.35, 1.00)$

Sample $U \sim \text{Unif}(0, 1)$. Choose the corresponding x-value, i.e., $X = F^{-1}(U)$. For example, if $U = 0.46$ then we have that $X = 4.2$. If $U = 0.20$ then $X = -2$. In this way we obtain a discrete random variable

□

3.7 Expected Value

The expected value of a random variable is defined formally below, but informally you can think of it as the "most likely outcome" or slightly closer to the truth as the amount that you will end up with if you perform an experiment over and over again. For example, if we play a game where I get \$1 from you for every heads I flip and you get \$1 from me for every tails I flip (for a fair coin) then the expected value of the game (for either of us) is \$0. But if we play that we each toss a coin and I only win if they are *both* heads, then the expected value of the game is $-50c$ for me and $50c$ for you. Since our respective expected values sum to zero this is called a *zero sum game*.

Definition: The *expected value* (or *mean*) of a RV X is

$$\mathrm{E}[X] \equiv \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x)\, dx & \text{if } X \text{ is continuous} \end{cases} = \int_{\mathbb{R}} x\, dF(x).$$

Example: Suppose that $X \sim \text{Bernoulli}(p)$. Then

$$X = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \ (= q) \end{cases}$$

and we have $\mathrm{E}[X] = \sum_x x f(x) = p$.

□

Note: This is like saying if you have an unfair coin which flips heads with probability p and tails with probability $1-p$, then assuming $X = 1$ is heads and $X = 0$ is tails you get that the expected value of the coin toss is $E(X) = (1)p + (0)(1-p) = p$.

Example: Suppose that $X \sim \text{Uniform}(a, b)$. Then

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

and we have $\mathrm{E}[X] = \int_{\mathbb{R}} x f(x)\, dx = (a+b)/2$.

□

Note: Again, this makes sense if you think about it because the expected "location" of the random variable X which is uniform over the interval $[a, b]$ is right in the middle, at the average of the two values of the endpoints, at $(a+b)/2$.

Example: Suppose that $X \sim \text{Exponential}(\lambda)$. Then

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

and we have (after integration by parts and L'Hopitals Rule)

$$\mathrm{E}[X] \;=\; \int_{\mathbb{R}} x f(x)\, dx \;=\; \int_0^\infty x\lambda e^{-\lambda x}\, dx \;=\; \frac{1}{\lambda}.$$

$\square$

Def/Thm: (the "Law of the Unconscious Statistician" or "LOTUS" aka "The Flower Rule"): Suppose that $h(X)$ is some function of the RV X. Then

$$\mathrm{E}[h(X)] = \begin{cases} \sum_x h(x)f(x) & \text{if } X \text{ is disc} \\ \int_{\mathbb{R}} h(x)f(x)\,dx & \text{if } X \text{ is cts} \end{cases} \;=\; \int_{\mathbb{R}} h(x)\, dF(x).$$

The function $h(X)$ can be anything "nice", e.g., $h(X) = X^2$ or $1/X$ or $\sin(X)$ or $\ln(X)$.

Example: Suppose X is the following discrete RV:

x	2	3	4
$f(x)$	0.3	0.6	0.1

Then $\mathrm{E}[X^3] = \sum_x x^3 f(x) = 8(0.3) + 27(0.6) + 64(0.1) = 25$.

$\square$

Example: Suppose $X \sim \mathrm{Unif}(0,2)$. Then

$$\mathrm{E}[X^n] \;=\; \int_{\mathbb{R}} x^n f(x)\, dx \;=\; 2^n/(n+1).$$

$\square$

3.8 Workshop 6: Random variables

Discrete RV Problems

1. A die is rolled 5 times. Let X denote the number of times that you see a 4, 5, or 6.

   (a) What's the distribution of X?

   Solution: $X \sim \text{Bin}(5, 1/2)$. □

   (b) Find $P(X = 4)$.

   Solution:

   $$P(X = 4) \;=\; \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{5-4} \;=\; 5/32. \quad \square$$

2. Suppose $X \sim \text{Pois}(2)$. Find $P(X > 3)$.

   Solution:

   $$P(X > 3) \;=\; 1 - P(X \leq 3) \;=\; 1 - \sum_{k=0}^{3} e^{-2}2^k/k! \;=\; 0.1429. \quad \square$$

3. Suppose X has the following discrete distribution.

x	-1	0	2	3
$P(X = x)$	0.2c	0.3	0.2	0.1

   (a) Find the value of c that will make the p.m.f. sum to 1.

   Solution: Note that

   $$1 \;=\; \sum_x P(X = x) \;=\; 0.2c + 0.3 + 0.2 + 0.1.$$

   This implies that $c = 2$. □

   (b) Find the c.d.f. $F(x)$ for all x.

Solution: We have

$$F(x) \;=\; P(X \le x) \;=\; \begin{cases} 0 & \text{if } x < -1 \\ 0.4 & \text{if } -1 \le x < 0 \\ 0.7 & \text{if } 0 \le x < 2 \\ 0.9 & \text{if } 2 \le x < 3 \\ 1.0 & \text{if } x \ge 3 \end{cases} \qquad \square$$

(c) Calculate E[X].

Solution:

$$\mathrm{E}[X] \;=\; \sum_x xP(X = x) \;=\; 0.3. \qquad \square$$

(d) Calculate Var(X).

Solution:

$$\mathrm{E}[X^2] \;=\; \sum_x x^2 P(X = x) \;=\; 2.1.$$

This implies that $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 2.01$. $\square$

(e) Calculate $P(1 \le X \le 2)$.

Solution: Since X can only equal $-1, 0, 2, 3$, we have $P(1 \le X \le 2) = P(X = 2) = 0.2$. $\square$

4. Suppose that X is the lifetime of a lightbulb and that $X \sim \mathrm{Exp}(2/\text{year})$.

(a) What's the probability that the bulb will survive at least a year, $P(X > 1)$?

Solution: If $X \sim \mathrm{Exp}(\lambda)$, we know from class that $P(X > t) = e^{-\lambda t}$. Thus, in this problem, we find that $P(X > 1) = e^{-2}$. $\square$

(b) Suppose the bulb has already survived a year. What's the probability that it will survive another year, i.e., $P(X > 2 \,|\, X > 1)$?

Solution: By conditional probability, we have

$$P(X > 2 \,|\, X > 1) \;=\; \frac{P(X > 2 \,\cap\, X > 1)}{P(X > 1)} \;=\; \frac{P(X > 2)}{P(X > 1)} \;=\; \frac{e^{-4}}{e^{-2}} \;=\; e^{-2}.$$

Note that this is the same answer as in (a), and is an example of what is called the memoryless property (which we will talk about later). $\square$

5. Suppose that X is continuous with p.d.f. $f(x) = cx^2$, $0 \le x \le 1$.

   (a) Find the value of c that will make the p.d.f. integrate to 1.

   Solution: Note that

   $$1 \;=\; \int_{\Re} f(x)\,dx \;=\; \int_0^1 cx^2\,dx \;=\; c/3.$$

   This implies that $c = 3$. $\square$

   (b) Find the c.d.f. $F(x)$ for all x.

   Solution:

   $$F(x) \;=\; P(X \le x) \;=\; \begin{cases} 0 & \text{if } x < 0 \\ x^3 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 1 \end{cases} \qquad \square$$

   (c) Calculate $\mathrm{E}[X]$.

   Solution:

   $$\mathrm{E}[X] \;=\; \int_{\Re} x f(x)\,dx \;=\; \int_0^1 3x^3\,dx = 3/4. \qquad \square$$

   (d) Calculate $\mathrm{Var}(X)$.

   Solution: Similarly,

   $$\mathrm{E}[X^2] \;=\; \int_{\Re} x^2 f(x)\,dx \;=\; \int_0^1 3x^4\,dx \;=\; 3/5.$$

   This implies that $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 3/80$. $\square$

   (e) Calculate $P(0 \le X \le 1/2)$.

   Solution: $P(0 \le X \le 1/2) = F(1/2) - F(0) = 1/8$. $\square$

6. Let $\mathrm{E}[X] = -4$, $\mathrm{Var}(X) = 5$, and $Z = -4X + 7$. Find $\mathrm{E}[-3Z]$ and $\mathrm{Var}(-3Z)$.

   Solution: We have

   $$\mathrm{E}[-3Z] \;=\; \mathrm{E}[12X - 21] \;=\; 12\mathrm{E}[X] - 21 \;=\; -69. \qquad \square$$

and

$$\text{Var}(-3Z) \;=\; \text{Var}(12X-21) \;=\; 144\text{Var}(X) \;=\; 720. \quad \Box$$

7. When a machine is adjusted properly, 50% of the items it produces are good and 50% are bad. However, the machine is *improperly* adjusted 10% of the time; in this case, 25% of the items it makes are good and 75% are bad. Suppose that 5 items produced by the machine are selected at random and inspected. If 4 of these items are good (and 1 is bad), what's the probability that the machine was adjusted properly at the time? **Hint:** Try Bayes Theorem using Binomial conditional probabilities.

   Solution: Let X be the number of good items (out of 5). Further, define the following events.

$$\begin{aligned} P &= \text{“machine is properly adjusted”, which implies } X \sim \text{Bin}(5, 1/2) \\ I &= \text{“machine is improperly adjusted”, which implies } X \sim \text{Bin}(5, 1/4) \end{aligned}$$

   Now Bayes implies that

$$\begin{aligned} P(P|X=4) &= \frac{P(X=4|P)P(P)}{P(X=4|P)P(P)+P(X=4|I)P(I)} \\ &= \frac{\binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1(0.9)}{\binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1(0.9)+\binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1(0.1)} \\ &= 0.9897. \quad \Box \end{aligned}$$

8. BONUS: Suppose that $X \sim \text{Unif}(-1, 6)$. Compare the upper bound on the probability $P(|X-\mu| \ge 1.5\sigma)$ obtained from Chebychev's inequality with the exact probability.

   Solution: By Chebychev, we get the following bound.

$$P(|X-\mu| \ge 1.5\sigma) \;\le \frac{1}{(1.5)^2} \;=\; 4/9. \quad \Box$$

   Now let's get the exact probability. First of all, recall that if $X \sim \text{U}(a, b)$, then $\text{E}[X] = (a+b)/2$ and $\text{Var}(X) = (a-b)^2/12$. Since $X \sim \text{U}(-1, 6)$, we have $\text{E}[X] = 5/2$ and $\sigma^2 = \text{Var}(X) = 49/12$. Thus, we can make the following exact calculations.

$$\begin{aligned}
P(|X-\mu| \geq 1.5\sigma) &= P\left(\left|X-\frac{5}{2}\right| \geq 1.5\cdot\sqrt{49/12}\right) \\
&= 1-P\left(\left|X-\frac{5}{2}\right| < 3.03\right) \\
&= 1-P\left(-3.03 < X-\frac{5}{2} < 3.03\right) \\
&= 1-P\left(-0.53 < X < 5.53\right) \\
&= 1-\int_{-0.53}^{5.53} f(x)\,dx \\
&= 1-\frac{1}{7}(5.53+0.53) \;=\; 0.1343. \qquad \square
\end{aligned}$$

9. What Zombies song is based on Ben E. King's "Stand By Me" (which was performed at the recent Royal Wedding)?

   Solution: "Time of the Season" $\square$

Continuous RV Problems

1. Suppose $X \sim \text{Unif}(1,3)$. Find the p.d.f. of $Z = e^X$.

   Hint: The c.d.f. of Z is

$$\begin{aligned}
G(z) &= P(Z \leq z) \\
&= P(e^X \leq z) \\
&= P(X \leq \ell\text{n}(z)) \\
&= \int_1^{\ell\text{n}(z)} f(x)\,dx \quad (\text{if } 1 \leq \ell\text{n}(z) \leq 3) \\
&= (\ell\text{n}(z)-1)/2.
\end{aligned}$$

   Now you can get the p.d.f.

$$g(z) \;=\; \frac{d}{dz}G(z) \;=\; \begin{cases} 0 & \text{if } z < e \text{ or } z > e^3 \\ \frac{1}{2z} & \text{if } e \leq z \leq e^3 \end{cases} \qquad \square$$

2. Suppose X has p.d.f. $f(x) = 2xe^{-x^2}$, $x \geq 0$. Find the distribution of $Z = X^2$.

Hint: The c.d.f. of Z is

$$\begin{aligned} G(z) &= P(Z \le z) \\ &= P(X^2 \le z) \\ &= P(-\sqrt{z} \le X \le \sqrt{z}) \\ &= P(0 \le X \le \sqrt{z}) \quad (\text{since } X \ge 0) \\ &= \int_0^{\sqrt{z}} 2xe^{-x^2}\,dx \\ &= 1 - e^{-z}. \end{aligned}$$

Thus, Z is Exp(1). $\square$

3. Suppose $X \sim \text{Unif}(1,3)$. Find the p.d.f. of $Z = e^X$.

Solution: The c.d.f. of Z is

$$\begin{aligned} G(z) &= P(Z \le z) \\ &= P(e^X \le z) \\ &= P(X \le \ell\text{n}(z)) \\ &= \int_1^{\ell\text{n}(z)} f(x)\,dx \quad (\text{if } 1 \le \ell\text{n}(z) \le 3) \\ &= (\ell\text{n}(z) - 1)/2. \end{aligned}$$

Now you can get the p.d.f.

$$g(z) = \frac{d}{dz}G(z) = \begin{cases} 0 & \text{if } z < e \text{ or } z > e^3 \\ \frac{1}{2z} & \text{if } e \le z \le e^3 \end{cases} \qquad \square$$

4. Suppose X has p.d.f. $f(x) = 2xe^{-x^2}$, $x \ge 0$. Find the distribution of $Z = X^2$.

Solution: The c.d.f. of Z is

$$\begin{aligned} G(z) &= P(Z \le z) \\ &= P(X^2 \le z) \\ &= P(-\sqrt{z} \le X \le \sqrt{z}) \\ &= P(0 \le X \le \sqrt{z}) \quad (\text{since } X \ge 0) \\ &= \int_0^{\sqrt{z}} 2xe^{-x^2}\,dx \\ &= 1 - e^{-z}. \end{aligned}$$

Thus, Z is Exp(1). $\square$

5. (Hines et al., 4–1). A refrigerator manufacturer subjects his finished products to a final inspection. Of interest are two categories of defects: scratches or flaws in the porcelain finish, and mechanical defects. The number of each type of defects is a random variable. The results of inspecting 50 refrigerators are shown in the following joint p.m.f. table, where X represents the occurrence of finish defects and Y represents the occurrence of mechanical defects.

$Y \backslash X$	0	1	2	3	4	5
0	11/50	4/50	2/50	1/50	1/50	1/50
1	8/50	3/50	2/50	1/50	1/50	
2	4/50	3/50	2/50	1/50		
3	3/50	1/50				
4	1/50					

(a) Find the marginal probability mass functions of X and Y.

Solution: Let's re-write the table, this time including the marginals.

$Y \backslash X$	0	1	2	3	4	5	$f_Y(y)$
0	11/50	4/50	2/50	1/50	1/50	1/50	20/50
1	8/50	3/50	2/50	1/50	1/50		15/50
2	4/50	3/50	2/50	1/50			10/50
3	3/50	1/50					4/50
4	1/50						1/50
$f_X(x)$	27/50	11/50	6/50	3/50	2/50	1/50	

□

(b) Find the conditional p.m.f. of mechanical defects, given that there are no finish defects.

Solution:

$$f(y|X=0) = \frac{f(0,y)}{f_X(0)} = \frac{f(0,y)}{27/50} = \begin{cases} 11/27 & \text{if } y=0 \\ 8/27 & \text{if } y=1 \\ 4/27 & \text{if } y=2 \\ 3/27 & \text{if } y=3 \\ 1/27 & \text{if } y=4 \\ 0 & \text{otherwise} \end{cases}$$

□

6. Suppose that $f(x,y) = cxy^2$ for $0 < x < y^2 < 1$ and $0 < y < 1$.

(a) Find c.

Solution:

$$1 \;=\; \int\int_{\mathbb{R}^2} f(x,y)\,dx\,dy \;=\; \int_0^1 \int_0^{y^2} cxy^2\,dx\,dy \;=\; c/14.$$

This immediately implies that $c = 14$. ☐

(b) Find the marginal p.d.f. of X, $f_X(x)$.

Solution:

$$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dy \;=\; \int_{\sqrt{x}}^1 14xy^2\,dy \;=\; \frac{14}{3}(x - x^{5/2}), \quad 0 < x < 1.$$ ☐

(c) Find the marginal p.d.f. of Y, $f_Y(y)$.

Solution:

$$f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx \;=\; \int_0^{y^2} 14xy^2\,dx \;=\; 7y^6, \quad 0 < y < 1.$$ ☐

(d) Find $\mathrm{E}[X]$.

Solution:

$$\mathrm{E}[X] \;=\; \int_{\mathbb{R}} x f_X(x)\,dx \;=\; \int_0^1 \frac{14}{3}(x^2 - x^{7/2})\,dx \;=\; \frac{14}{27}.$$ ☐

(e) Find $\mathrm{E}[Y]$.

Solution:

$$\mathrm{E}[Y] \;=\; \int_{\mathbb{R}} y f_Y(y)\,dy \;=\; \int_0^1 7y^7\,dy \;=\; \frac{7}{8}.$$ ☐

(f) Find the conditional p.d.f. of X given $Y = y$, $f(x|y)$.

Solution:

$$f(x|y) \;=\; \frac{f(x,y)}{f_Y(y)} \;=\; \frac{2x}{y^4},\; 0 < x < y^2 < 1.$$ ☐

7. **Mathemusical Bonus:** What is the largest prime number to be found in the lyrics of a song from the Top-40 era?

   Solution: Tommy Tutone's song "Jenny" mentions the prime number 8675309.
   www.youtube.com/watch?v=6WTdTwcmxyo $\square$

3.9 Moments and Standard Deviation

Definitions: $\mathrm{E}[X^n]$ is the nth *moment* of X.
$\mathrm{E}[(X - \mathrm{E}[X])^n]$ is the nth *central moment* of X.
$\mathrm{Var}(X) := \mathrm{E}[(X - \mathrm{E}[X])^2]$ is the *variance* of X.
The *standard deviation* of X is $\sqrt{\mathrm{Var}(X)}$.

Theorem: $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$

Example: Suppose $X \sim \mathrm{Bern}(p)$. Recall that $\mathrm{E}[X] = p$. Then

$$\mathrm{E}[X^2] = \sum_x x^2 f(x) \;=\; p \quad \text{and}$$

$$\mathrm{Var}(X) \;=\; \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \;=\; p(1-p).$$

□

Example: Suppose $X \sim \mathrm{Exp}(\lambda)$. By LOTUS,

$$\mathrm{E}[X^n] \;=\; \int_0^\infty x^n \lambda e^{-\lambda x}\, dx \;=\; n!/\lambda^n.$$

$$\mathrm{Var}(X) \;=\; \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \;=\; \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \;=\; 1/\lambda^2.$$

□

Theorem: $\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$ and $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

Example: If $X \sim \mathrm{Exp}(3)$, then

$$\mathrm{E}[-2X + 7] \;=\; -2\mathrm{E}[X] + 7 \;=\; -\frac{2}{3} + 7.$$

$$\mathrm{Var}(-2X + 7) \;=\; (-2)^2\mathrm{Var}(X) \;=\; \frac{4}{9}.$$

□

3.10 Moment Generating Functions

Definition: $M_X(t) := \mathrm{E}[e^{tX}]$ is the *moment generating function* (mgf) of the RV X. ($M_X(t)$ is a function of t, *not* of X!)

Example: : $X \sim \mathrm{Bern}(p)$. Then

$$M_X(t) \;=\; \mathrm{E}[e^{tX}] \;=\; \sum_x e^{tx} f(x) \;=\; e^{t\cdot 1}p + e^{t\cdot 0}q \;=\; pe^t + q. \quad \Box$$

Example: : $X \sim \mathrm{Exp}(\lambda)$. Then

$$M_X(t) \;=\; \int_{\Re} e^{tx} f(x)\,dx \;=\; \lambda \int_0^\infty e^{(t-\lambda)x}\,dx \;=\; \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \Box$$

Theorem: Under certain technical conditions,

$$\mathrm{E}[X^k] \;=\; \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}, \quad k = 1, 2, \ldots.$$

Thus, you can *generate* the moments of X by taking derivatives of the the mgf. (hmmm... what else were we talking about recently that you could figure out from taking a bunch of derivatives of something...?)

Example: : $X \sim \mathrm{Exp}(\lambda)$. Then $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda > t$. So

$$\mathrm{E}[X] \;=\; \frac{d}{dt} M_X(t)\bigg|_{t=0} \;=\; \frac{\lambda}{(\lambda - t)^2}\bigg|_{t=0} \;=\; 1/\lambda.$$

Further,

$$\mathrm{E}[X^2] \;=\; \frac{d^2}{dt^2} M_X(t)\bigg|_{t=0} \;=\; \frac{2\lambda}{(\lambda - t)^3}\bigg|_{t=0} \;=\; 2/\lambda^2.$$

Thus,

$$\mathrm{Var}(X) \;=\; \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \;=\; \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \;=\; 1/\lambda^2. \quad \Box$$

Moment generating functions have many other important uses, some of which we'll talk about in this course.

3.11 Functions of Random Variables

Problem: Suppose we have a RV X with pmf/pdf $f(x)$. Let $Y = h(X)$. Find $g(y)$, the pmf/pdf of Y.

Examples (can be proven):

If $X \sim \text{Nor}(0,1)$, then $Y = X^2 \sim \chi^2(1)$.

If $U \sim \text{Unif}(0,1)$, then $Y = -\frac{1}{\lambda}\ln(U) \sim \text{Exp}(\lambda)$.

□

Discrete Example: Let X denote the number of H's from two coin tosses. We want the pmf for $Y = X^3 - X$.

x	0	1	2
$f(x)$	1/4	1/2	1/4
$y = x^3 - x$	0	0	6

This implies that $g(0) = P(Y = 0) = P(X = 0 \text{ or } 1) = 3/4$ and $g(6) = P(Y = 6) = 1/4$. In other words,

$$g(y) = \begin{cases} 3/4 & \text{if } y = 0 \\ 1/4 & \text{if } y = 6 \end{cases}.$$

□

Continuous Example: Suppose X has pdf $f(x) = |x|$, $-1 \le x \le 1$. Let's find the pdf of $Y = X^2$.

First of all, the cdf of Y is

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} |x|\,dx \;=\; y, \quad 0 < y < 1. \end{aligned}$$

Thus, the pdf of Y is $g(y) = G'(y) = 1$, $0 < y < 1$, indicating that $Y \sim \text{Unif}(0,1)$.

□

3.12 Jointly Distributed Random Variables

Consider two random variables interacting together — think height and weight.

Definition: The *joint cdf* of X and Y is

$$F(x,y) \;:=\; P(X \le x, Y \le y), \quad \text{for all } x, y.$$

Remark: The *marginal cdf* of X is $F_X(x) = F(x, \infty)$. (We use the X subscript to remind us that it's just the cdf of X all by itself.) Similarly, the *marginal cdf* of Y is $F_Y(y) = F(\infty, y)$.

Definition: If X and Y are discrete, then the *joint pmf* of X and Y is $f(x,y) \equiv P(X = x, Y = y)$. Note that $\sum_x \sum_y f(x,y) = 1$.

Remark: The *marginal pmf* of X is

$$f_X(x) \;=\; P(X = x) \;=\; \sum_y f(x,y).$$

The *marginal pmf* of Y is

$$f_Y(y) \;=\; P(Y = y) \;=\; \sum_x f(x,y).$$

Example: The following table gives the joint pmf $f(x,y)$, along with the accompanying marginals.

$f(x,y)$	$X = 2$	$X = 3$	$X = 4$	$f_Y(y)$
$Y = 4$	0.3	0.2	0.1	0.6
$Y = 6$	0.1	0.2	0.1	0.4
$f_X(x)$	0.4	0.4	0.2	1

□

Definition: If X and Y are continuous, then the *joint pdf* of X and Y is $f(x,y) \equiv \frac{\partial^2}{\partial x \partial y} F(x,y)$. Note that $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\,dx\,dy = 1$.

Remark: The *marginal pdf's* of X and Y are

$$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dy \quad \text{and} \quad f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx.$$

Example: Suppose the joint pdf is

$$f(x,y) \;=\; \frac{21}{4} x^2 y, \quad x^2 \le y \le 1.$$

Then the marginal pdf's are:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4}x^2 y\,dy = \frac{21}{8}x^2(1-x^4),\ -1 \le x \le 1$$

and

$$f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx \;=\; \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2 y\,dx \;=\; \frac{7}{2}y^{5/2},\quad 0 \le y \le 1. \quad \Box$$

$\Box$

3.13 Independent Random Variables

Definition: X and Y are *independent* RV's if

$$f(x,y) \;=\; f_X(x) f_Y(y) \quad \text{for all } x, y.$$

Theorem: X and Y are indep if you can write their joint pdf as $f(x,y) = a(x)b(y)$ for some functions $a(x)$ and $b(y)$, and x and y don't have funny limits (their domains do not depend on each other).

Examples: If $f(x,y) = cxy$ for $0 \le x \le 2$, $0 \le y \le 3$, then X and Y are independent.

If $f(x,y) = \frac{21}{4}x^2 y$ for $x^2 \le y \le 1$, then X and Y are *not* independent.

If $f(x,y) = c/(x+y)$ for $1 \le x \le 2$, $1 \le y \le 3$, then X and Y are *not* independent.

$\Box$

Definition: The *conditional pdf* (or *pmf*) of Y given $X = x$ is $f(y|x) \equiv f(x,y)/f_X(x)$ (assuming $f_X(x) > 0$).

This is a legit pmf/pdf. For example, in the continuous case, $\int_{\mathbf{R}} f(y|x)\,dy = 1$, for any x.

Example: Suppose $f(x,y) = \frac{21}{4}x^2 y$ for $x^2 \le y \le 1$. Then

$$f(y|x) \;=\; \frac{f(x,y)}{f_X(x)} \;=\; \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)} \;=\; \frac{2y}{1-x^4},\quad x^2 \le y \le 1. \quad \Box$$

Theorem: If X and Y are indep, then $f(y|x) = f_Y(y)$ for all x, y.

Proof: By definition of conditional and independence,

$$f(y|x) \;=\; \frac{f(x,y)}{f_X(x)} \;=\; \frac{f_X(x) f_Y(y)}{f_X(x)}. \quad \Box$$

Definition: The *conditional expectation* of Y given $X = x$ is

$$\mathrm{E}[Y|X = x] \equiv \begin{cases} \sum_y y f(y|x) & \text{discrete} \\ \int_{\mathbb{R}} y f(y|x)\, dy & \text{continuous} \end{cases}$$

Example: : The expected weight of a person who is 7 feet tall ($\mathrm{E}[Y|X = 7]$) will probably be greater than that of a random person from the entire population ($\mathrm{E}[Y]$).

Example: : $f(x, y) = \frac{21}{4}x^2 y$, if $x^2 \le y \le 1$. Then

$$\mathrm{E}[Y|x] = \int_{\mathbb{R}} y f(y|x)\, dy = \int_{x^2}^{1} \frac{2y^2}{1 - x^4}\, dy = \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4}. \quad \Box$$

3.14 Covariance and Correlation

"Definition" (two-dimensional LOTUS): Suppose that $h(X,Y)$ is some function of the RV's X and Y. Then

$$\mathrm{E}[h(X,Y)] \;=\; \begin{cases} \sum_x \sum_y h(x,y)f(x,y) & \text{if } (X,Y) \text{ is discrete} \\ \int_{\mathbb{R}}\int_{\mathbb{R}} h(x,y)f(x,y)\,dx\,dy & \text{if } (X,Y) \text{ is continuous} \end{cases}$$

Theorem: Whether or not X and Y are independent, we have $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$.

Theorem: If X and Y are *independent*, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Definition: $X_1, \ldots, X_n$ form a *random sample* from $f(x)$ if (i) $X_1, \ldots, X_n$ are independent, and (ii) each X_i has the same pdf (or pmf) $f(x)$.

Notation: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$. (The term "iid" reads *independent and identically distributed.*)

Example: If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ and the *sample mean* $\bar{X}_n \equiv \sum_{i=1}^n X_i/n$, then $\mathrm{E}[\bar{X}_n] = \mathrm{E}[X_i]$ and $\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(X_i)/n$. Thus, the variance *decreases* as n increases. $\Box$

Definition: The *covariance* between X and Y is

$$\mathrm{Cov}(X,Y) \;\equiv\; \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] \;=\; \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

Note that $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

Theorem: If X and Y are independent RV's, then $\mathrm{Cov}(X,Y) = 0$.

Remark: $\mathrm{Cov}(X,Y) = 0$ doesn't mean X and Y are independent!

Example: Suppose $X \sim \mathrm{Unif}(-1,1)$ and $Y = X^2$. Then X and Y are clearly dependent. However,

$$\mathrm{Cov}(X,Y) \;=\; \mathrm{E}[X^3] - \mathrm{E}[X]\mathrm{E}[X^2] \;=\; \mathrm{E}[X^3] \;=\; \int_{-1}^{1} \frac{x^3}{2}\,dx \;=\; 0.$$

$\Box$

Theorem: $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X,Y)$.

Theorem: Whether or not X and Y are independent,

$$\mathrm{Var}(X+Y) \;=\; \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

and

$$\mathrm{Var}(X-Y) \;=\; \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X,Y).$$

Definition: The *correlation* between X and Y is

$$\rho \equiv \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Theorem: $-1 \le \rho \le 1$.

Example: Consider the following joint pmf.

$f(x,y)$	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=40$	0.00	0.20	0.10	0.3
$Y=50$	0.15	0.10	0.05	0.3
$Y=60$	0.30	0.00	0.10	0.4
$f_X(x)$	0.45	0.30	0.25	1

$\mathrm{E}[X]=2.8$, $\mathrm{Var}(X)=0.66$, $\mathrm{E}[Y]=51$, $\mathrm{Var}(Y)=69$,

$$\mathrm{E}[XY] = \sum_x \sum_y xyf(x,y) = 140,$$

and

$$\rho = \frac{\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = -0.415. \quad \square$$

Portfolio Example: Consider two assets, S_1 and S_2, with expected returns $\mathrm{E}[S_1]=\mu_1$ and $\mathrm{E}[S_2]=\mu_2$, and variabilities $\mathrm{Var}(S_1)=\sigma_1^2$, $\mathrm{Var}(S_2)=\sigma_2^2$, and $\mathrm{Cov}(S_1,S_2)=\sigma_{12}$.

Define a *portfolio* $P = wS_1 + (1-w)S_2$, where $w \in [0,1]$. Then

$$\mathrm{E}[P] = w\mu_1 + (1-w)\mu_2$$
$$\mathrm{Var}(P) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\sigma_{12}.$$

Setting $\frac{d}{dw}\mathrm{Var}(P) = 0$, we obtain the critical point that (hopefully) minimizes the variance of the portfolio,

$$w = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}. \quad \square$$

3.15 Summary of Distributions

1. $X \sim \mathrm{Bernoulli}(p)$.

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p\ (=q) & \text{if } x = 0 \end{cases}$$

$\mathrm{E}[X]=p$, $\mathrm{Var}(X)=pq$, $M_X(t) = pe^t + q$.

2. $Y \sim \text{Binomial}(n, p)$.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$ (i.e., *Bernoulli(p) trials*), then $Y = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$.

$$f(y) = \binom{n}{y} p^y q^{n-y}, \quad y = 0, 1, \ldots, n.$$

$\text{E}[Y] = np$, $\text{Var}(Y) = npq$, $M_Y(t) = (pe^t + q)^n$.

3. $X \sim \text{Geometric}(p)$

is the number of Bern(p) trials until a success occurs. For example, "FFFS" implies that $X = 4$.

$$f(x) = q^{x-1}p, \quad x = 1, 2, \ldots.$$

$\text{E}[X] = 1/p$, $\text{Var}(X) = q/p^2$, $M_X(t) = pe^t/(1 - qe^t)$.

4. $Y \sim \text{NegBin}(r, p)$

is the sum of r iid Geom(p) RV's, i.e., the time until the rth success occurs. For example, "FFFSSFS" implies that $\text{NegBin}(3, p) = 7$.

$$f(y) = \binom{y-1}{r-1} q^{y-r} p^r, \quad y = r, r+1, \ldots.$$

$\text{E}[Y] = r/p$, $\text{Var}(Y) = qr/p^2$.

5. $X \sim \text{Poisson}(\lambda)$.

Definition: A *counting process* $N(t)$ tallies the number of "arrivals" observed in $[0, t]$. A *Poisson process* is a counting process satisfying the following.

i. Arrivals occur one-at-a-time at rate λ (e.g., $\lambda = 4$ customers/hr)

ii. Independent increments, i.e., the numbers of arrivals in disjoint time intervals are independent.

iii. Stationary increments, i.e., the distribution of the number of arrivals in $[s, s+t]$ only depends on t.

Then $X \sim \text{Pois}(\lambda)$ is the number of arrivals that a Poisson process experiences in one time unit, i.e., $N(1)$.

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots.$$

$\text{E}[X] = \lambda = \text{Var}(X)$, $M_X(t) = e^{\lambda(e^t - 1)}$.

And the continuous distributions we discussed

1. $X \sim \text{Uniform}(a,b)$.

$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, $\text{E}[X] = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$, $M_X(t) = (e^{tb} - e^{ta})/(tb - ta)$.

2. $X \sim \text{Exponential}(\lambda)$. $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, $\text{E}[X] = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$, $M_X(t) = \lambda/(\lambda - t)$ for $t < \lambda$.

Theorem: The exponential distribution has the *memoryless property* (and is the only continuous distribution with this property), i.e., for $s,t > 0$, $P(X > s+t | X > s) = P(X > t)$.

Example: Suppose $X \sim \text{Exp}(\lambda = 1/100)$. Then

$$P(X > 200 | X > 50) \;=\; P(X > 150) \;=\; e^{-\lambda t} \;=\; e^{-150/100}. \quad \Box$$

3. $X \sim \text{Gamma}(\alpha, \lambda)$.

$$f(x) \;=\; \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x \geq 0,$$

where the gamma function is

$$\Gamma(\alpha) \;\equiv\; \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

$\text{E}[X] = \alpha/\lambda$, $\text{Var}(X) = \alpha/\lambda^2$, $M_X(t) = \Big[\lambda/(\lambda - t)\Big]^\alpha$ for $t < \lambda$.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, then $Y \equiv \sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$. The $\text{Gamma}(n,\lambda)$ is also called the $\text{Erlang}_n(\lambda)$. It has cdf

$$F_Y(y) \;=\; 1 - e^{-\lambda y} \sum_{j=0}^{n-1} \frac{(\lambda y)^j}{j!}, \quad y \geq 0.$$

4. $X \sim \text{Triangular}(a,b,c)$.

Good for modeling things with limited data — a is the smallest possible value, b is the "most likely," and c is the largest.

$$f(x) \;=\; \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a < x \leq b \\ \frac{2(c-x)}{(c-b)(c-a)} & \text{if } b < x \leq c \\ 0 & \text{otherwise} \end{cases}.$$

$\text{E}[X] = (a+b+c)/3$.

5. $X \sim \text{Beta}(a,b)$. $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ for $0 \le x \le 1$ and $a,b>0$.

$$\text{E}[X] \;=\; \frac{a}{a+b} \quad \text{and} \quad \text{Var}(X) \;=\; \frac{ab}{(a+b)^2(a+b+1)}.$$

6. $X \sim \text{Normal}(\mu, \sigma^2)$. Most important distribution.

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

$\text{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$, and $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

Theorem: If $X \sim \text{Nor}(\mu, \sigma^2)$, then $aX + b \sim \text{Nor}(a\mu + b, a^2\sigma^2)$.

Corollary: If $X \sim \text{Nor}(\mu, \sigma^2)$, then $Z \equiv \frac{X-\mu}{\sigma} \sim \text{Nor}(0,1)$, the *standard normal distribution*, with pdf $\phi(z) \equiv \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ and cdf $\Phi(z)$, which is tabled. E.g., $\Phi(1.96) \doteq 0.975$.

Theorem: If X_1 and X_2 are *independent* with $X_i \sim \text{Nor}(\mu_i, \sigma_i^2)$, $i = 1,2$, then $X_1 + X_2 \sim \text{Nor}(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)$.

Example: Suppose $X \sim \text{Nor}(3,4)$, $Y \sim \text{Nor}(4,6)$, and X and Y are independent. Then $2X - 3Y + 1 \sim \text{Nor}(-5, 70)$.

$\square$

3.16 Limit Theorems

Corollary (of a previous theorem): If $X_1, \ldots, X_n$ are iid $\text{Nor}(\mu, \sigma^2)$, then the sample mean $\bar{X}_n \sim \text{Nor}(\mu, \sigma^2/n)$.

This is a special case of the *Law of Large Numbers*, which says that $\bar{X}_n$ approximates μ well as n becomes large.

Definition: The sequence of RV's $Y_1, Y_2, \ldots$ with respective cdf's $F_{Y_1}(y), F_{Y_2}(y), \ldots$ *converges in distribution* to the RV Y having cdf $F_Y(y)$ if $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ for all y belonging to the continuity set of Y. Notation: $Y_n \xrightarrow{d} Y$.

Idea: If $Y_n \xrightarrow{d} Y$ and n is large, then you ought to be able to approximate the distribution of Y_n by the limit distribution of Y.

Central Limit Theorem: If $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$ with mean μ and variance σ^2, then

$$Z_n \equiv \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \text{Nor}(0,1).$$

Thus, the cdf of Z_n approaches $\Phi(z)$ as n increases.

The CLT is the most-important theorem in the universe.

The CLT usually works well if the pmf/pdf is fairly symmetric and $n \geq 15$.

We will eventually look at more-general versions of the CLT described above.

Example: If $X_1, X_2, \ldots, X_{100} \stackrel{\text{iid}}{\sim} \text{Exp}(1)$ (so $\mu = \sigma^2 = 1$), then

$$\begin{aligned}
&P\Big(90 \leq \sum_{i=1}^{100} X_i \leq 110\Big) \\
&\quad = P\Big(\frac{90-100}{\sqrt{100}} \leq Z_{100} \leq \frac{110-100}{\sqrt{100}}\Big) \\
&\quad \approx P(-1 \leq \text{Nor}(0,1) \leq 1) = 0.6827.
\end{aligned}$$

By the way, since $\sum_{i=1}^{100} X_i \sim \text{Erlang}_{k=100}(\lambda = 1)$, we can use the cdf (which may be tedious) or software such as Minitab to obtain the *exact* value of $P(90 \leq \sum_{i=1}^{100} X_i \leq 110) = 0.6835$.

Wow! The CLT and exact answers match nicely! $\Box$

3.17 Statistics and Estimation

Definition: A *statistic* is a function of the observations $X_1, \ldots, X_n$, and not explicitly dependent on any unknown parameters.

Examples of statistics: $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

Statistics are *random variables*. If we take two different samples, we'd expect to get two different values of a statistic.

A statistic is usually used to estimate some unknown *parameter* from the underlying probability distribution of the X_i's.

Examples of parameters: μ, σ^2.

Let $X_1, \ldots, X_n$ be iid RV's and let $T(\mathbf{X}) \equiv T(X_1, \ldots, X_n)$ be a statistic based on the X_i's. Suppose we use $T(\mathbf{X})$ to estimate some unknown parameter θ. Then $T(\mathbf{X})$ is called a *point estimator* for θ.

Examples: $\bar{X}$ is usually a point estimator for the mean $\mu = \mathrm{E}[X_i]$, and S^2 is often a point estimator for the variance $\sigma^2 = \mathrm{Var}(X_i)$.

It would be nice if $T(\mathbf{X})$ had certain properties:

* Its expected value should equal the parameter it's trying to estimate.

* It should have low variance.

3.18 Unbiased Estimation

Definition: $T(\mathbf{X})$ is *unbiased* for θ if $\mathrm{E}[T(\mathbf{X})] = \theta$.

Example: /**Theorem**: Suppose $X_1, \ldots, X_n$ are iid anything with mean μ. Then

$$\mathrm{E}[\bar{X}] \;=\; \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] \;=\; \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i] \;=\; \mathrm{E}[X_i] \;=\; \mu.$$

So $\bar{X}$ is always unbiased for μ. That's why $\bar{X}$ is called the *sample mean*.

Baby Example: In particular, suppose $X_1, \ldots, X_n$ are iid Exp(λ). Then $\bar{X}$ is unbiased for $\mu = \mathrm{E}[X_i] = 1/\lambda$.

But be careful.... $1/\bar{X}$ is *biased* for λ in this exponential case, i.e., $\mathrm{E}[1/\bar{X}] \neq 1/\mathrm{E}[\bar{X}] = \lambda$.

Example: /**Theorem**: Suppose $X_1, \ldots, X_n$ are iid anything with mean μ and variance σ^2. Then

$$\mathrm{E}[S^2] \;=\; \mathrm{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] \;=\; \mathrm{Var}(X_i) \;=\; \sigma^2.$$

Thus, S^2 is always unbiased for σ^2. This is why S^2 is called the *sample variance.*

Baby Example: Suppose $X_1, \ldots, X_n$ are iid Exp(λ). Then S^2 is unbiased for $\mathrm{Var}(X_i) = 1/\lambda^2$.

3.19 Distributional Results and Confidence Intervals

There are a number of distributions (including the normal) that come up in statistical sampling problems. Here are a few:

Definitions: If $Z_1, Z_2, \ldots, Z_k$ are iid Nor(0,1), then $Y = \sum_{i=1}^k Z_i^2$ has the χ^2 *distribution with k degrees of freedom (df)*. Notation: $Y \sim \chi^2(k)$. Note that $\mathrm{E}[Y] = k$ and $\mathrm{Var}(Y) = 2k$.

If $Z \sim \mathrm{Nor}(0,1)$, $Y \sim \chi^2(k)$, and Z and Y are independent, then $T = Z/\sqrt{Y/k}$ has the *Student t distribution with k df.* Notation: $T \sim t(k)$. Note that the $t(1)$ is the *Cauchy* distribution.

If $Y_1 \sim \chi^2(m)$, $Y_2 \sim \chi^2(n)$, and Y_1 and Y_2 are independent, then $F = (Y_1/m)/(Y_2/n)$ has the *F distribution with m and n df.* Notation: $F \sim F(m,n)$.
How (and why) would one use the above facts? Because they can be used to construct *confidence intervals* (CIs) for μ and σ^2 under a variety of assumptions.

A $100(1-\alpha)\%$ two-sided CI for an unknown parameter θ is a random interval $[L, U]$ such that $P(L \le \theta \le U) = 1 - \alpha$.

Here are some examples / theorems, all of which assume that the X_i's are iid normal...

Example: If σ^2 is *known*, then a $100(1-\alpha)\%$ CI for μ is

$$\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}},$$

where z_γ is the $1-\gamma$ quantile of the standard normal distribution, i.e., $z_\gamma \equiv \Phi^{-1}(1-\gamma)$.

Example: If σ^2 is *unknown*, then a $100(1-\alpha)\%$ CI for μ is

$$\bar{X}_n - t_{\alpha/2,n-1}\sqrt{\frac{S^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + t_{\alpha/2,n-1}\sqrt{\frac{S^2}{n}},$$

where $t_{\gamma,\nu}$ is the $1-\gamma$ quantile of the $t(\nu)$ distribution.

Example: A $100(1-\alpha)\%$ CI for σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \;\le\; \sigma^2 \;\le\; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}},$$

where $\chi^2_{\gamma,\nu}$ is the $1-\gamma$ quantile of the $\chi^2(\nu)$ distribution.

Exercise: Here are 20 residual flame times (in sec.) of treated specimens of children's nightwear. (Don't worry — children were not in the nightwear when the clothing was set on fire.)

9.85	9.93	9.75	9.77	9.67
9.87	9.67	9.94	9.85	9.75
9.83	9.92	9.74	9.99	9.88
9.95	9.95	9.93	9.92	9.89

Let's get a 95% CI for the mean residual flame time.

After a little algebra, we get

$$\bar{X} \;=\; 9.8525 \quad \text{and} \quad S \;=\; 0.0965.$$

Further, you can use the Excel function `t.inv(0.975,19)` to get $t_{\alpha/2,n-1} = t_{0.025,19} = 2.093$.

Then the half-length of the CI is

$$H \;=\; t_{\alpha/2,n-1}\sqrt{S^2/n} \;=\; \frac{(2.093)(0.0965)}{\sqrt{20}} \;=\; 0.0451.$$

Thus, the CI is $\mu \in \bar{X} \pm H$, or $9.8074 \le \mu \le 9.8976$. $\Box$

3.20 Workshop 7: Advanced probability and statistics

1. Suppose that $f(x,y) = 14xy^2$ for $0 < x < y^2 < 1$ and $0 < y < 1$.

   (a) Find the marginal p.d.f. of X, $f_X(x)$.

   Solution:

   $$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dy \;=\; \int_{\sqrt{x}}^{1} 14xy^2\,dy \;=\; \frac{14}{3}(x - x^{5/2}), \quad 0 < x < 1. \qquad \square$$

   (b) Find the marginal p.d.f. of Y, $f_Y(y)$.

   Solution:

   $$f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx \;=\; \int_0^{y^2} 14xy^2\,dx \;=\; 7y^6, \quad 0 < y < 1. \qquad \square$$

   (c) Find $\mathrm{E}[X]$.

   Solution:

   $$\mathrm{E}[X] \;=\; \int_{\mathbb{R}} x f_X(x)\,dx \;=\; \int_0^1 \frac{14}{3}(x^2 - x^{7/2})\,dx \;=\; \frac{14}{27}. \qquad \square$$

   (d) Find the conditional p.d.f. of X given $Y = y$, $f(x|y)$.

   Solution:

   $$f(x|y) \;=\; \frac{f(x,y)}{f_Y(y)} \;=\; \frac{2x}{y^4},\ 0 < x < y^2 < 1. \qquad \square$$

   (e) Find the conditional expectation, $\mathrm{E}[X|y]$.

   Solution:

   $$\mathrm{E}[X|y] \;=\; \int_{\mathbb{R}} x f(x|y)\,dx \;=\; \int_0^{y^2} \frac{2x}{y^4}\,dx \;=\; \frac{2y^2}{3}. \qquad \square$$

   (f) Find the "double" conditional expectation, $\mathrm{E}[\mathrm{E}[X|Y]]$.

   Solution:

   $$\mathrm{E}[\mathrm{E}[X|Y]] \;=\; \int_{\mathbb{R}} \mathrm{E}[X|y] f_Y(y)\,dy \;=\; \int_0^1 \frac{2y^2}{3} 7y^6\,dy \;=\; \frac{14}{27}. \qquad \square$$

2. (Hines et al., 4–8.) Consider the probability distribution of the discrete random vector (X, Y), where X represents the number of orders for aspirin in August in the neighborhood drugstore and Y represents the number of orders in September. The joint distribution is shown in the following table.

$Y \backslash X$	51	52	53	54	55
51	0.06	0.05	0.05	0.01	0.01
52	0.07	0.05	0.01	0.01	0.01
53	0.05	0.10	0.10	0.05	0.05
54	0.05	0.02	0.01	0.01	0.03
55	0.05	0.06	0.05	0.01	0.03

(a) Find the marginal distributions.

Solution: After we add up the usual stuff, we get the following marginals:

z	51	52	53	54	55
$f_X(z)$	0.28	0.28	0.22	0.09	0.13
$f_Y(z)$	0.18	0.15	0.35	0.12	0.20

□

(b) Find the expected sales in September, given that sales in August were either 51, 52, 53, 54, or 55, respectively.

Solution:

$$\mathrm{E}[Y|X=x] \;=\; \sum_y y f_{Y|X}(y|x) \;=\; \frac{1}{f_X(x)} \sum_y y f(x,y).$$

For example, we get

$$\mathrm{E}[Y|X=51] \;=\; \frac{1}{0.28}\Big[(51)(0.06)+(52)(0.07)+\cdots+(55)(0.05)\Big] \;=\; 52.86.$$

Similarly, we get the following table.

x	51	52	53	54	55
$\mathrm{E}[Y\|X=x]$	52.86	52.96	53	53	53.46

□

3. (Hines et al., 4–9). Assume that X and Y are coded scores of two intelligence tests, and the p.d.f. of (X, Y) is given by

$$f(x,y) \;=\; \begin{cases} 6x^2y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value of the score on test #2 given the score on test #1.

Solution: Denote by X and Y the scores on tests #1 and #2, respectively. We want to eventually find $\text{E}[Y|X = x]$.

The marginal p.d.f. of X is

$$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dt \;=\; \int_0^1 6x^2y\,dy \;=\; 3x^2, \quad 0 \le x \le 1.$$

The conditional p.d.f. of $Y|X = x$ is

$$f(y|x) \;=\; \frac{f(x,y)}{f_X(x)} \;=\; \frac{6x^2y}{3x^2} \;=\; 2y, \quad 0 \le y \le 1.$$

The required conditional expectation is

$$\text{E}[Y|X = x] \;=\; \int_{\mathbb{R}} yf(y|x)\,dy \;=\; \int_0^1 2y^2\,dy \;=\; 2/3. \quad \Box$$

Note that this answer doesn't depend on x! This is because X and Y are *independent* (Why?) $\Box$

4. (Hines et al., 4–31.) Given the following joint p.d.f.'s, determine whether or not X and Y are independent.

   (a) $g(x,y) = 4xye^{-(x^2+y^2)}$, $x > 0$, $y > 0$.

   Solution: Since (i) there are no funny limits and (ii) you can factor $g(x,y) = (4xe^{-x^2})(ye^{-y^2})$, we see that X and Y are independent. $\Box$

   (b) $f(x,y) = 3x^2y^{-3}$, $0 < x < y < 1$.

   Solution: Funny limits imply *not* independent. $\Box$

   (c) $f(x,y) = 6(1 + x + y)^{-4}$, $x > 0$, $y > 0$.

   Solution: Can't factor $f(x,y) = g(x)h(y)$ implies *not* independent. $\Box$

5. (Hines et al., 4–19.) Let X and Y have joint p.d.f. $f(x,y) = 2$, $0 < x < y < 1$. Find the correlation between X and Y.

Solution: I won't go through all of the tedious calculations, but here are the highlights.

$$f_X(x) \;=\; \int_x^1 2\,dy \;=\; 2(1-x), \quad 0<x<1$$

and

$$f_Y(y) \;=\; \int_0^y 2\,dx \;=\; 2y, \quad 0<y<1.$$

Then you get (in the usual way)

$$\mathrm{E}[X] \;=\; 1/3, \quad \mathrm{Var}(X) \;=\; 1/18, \quad \mathrm{E}[Y] \;=\; 2/3, \quad \mathrm{Var}(Y) \;=\; 1/18.$$

Further,

$$\mathrm{E}[XY] \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy \;=\; \int_0^1\int_0^y 2xy\,dx\,dy \;=\; 1/4.$$

This finally implies that

$$\rho \;=\; \frac{\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \;=\; 0.5. \quad \Box$$

6. (Hines et al., 4–21). Consider the data from Hines et al., 4–1, reproduced below.

$Y \backslash X$	0	1	2	3	4	5
0	11/50	4/50	2/50	1/50	1/50	1/50
1	8/50	3/50	2/50	1/50	1/50	
2	4/50	3/50	2/50	1/50		
3	3/50	1/50				
4	1/50					

Are X and Y independent? Find the correlation.

Solution: After the usual manipulations, get $\rho = -0.1355$. So X and Y are *not* independent. $\Box$

7. Let $\mathrm{Var}(X) = \mathrm{Var}(Y) = 20$, $\mathrm{Var}(Z) = 30$, $\mathrm{Cov}(X,Y) = 2$, $\mathrm{Cov}(X,Z) = -3$, and $\mathrm{Cov}(Y,Z) = -4$. Find $\mathrm{Corr}(X,Z)$ and $\mathrm{Var}(X-2Y+5Z)$.

Solution:

$$\mathrm{Corr}(X,Z) \;=\; \frac{\mathrm{Cov}(X,Z)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Z)}} \;=\; -0.1225$$

and

$$\begin{aligned}\mathrm{Var}(X-2Y+5Z) &= \mathrm{Var}(X)+4\mathrm{Var}(Y)+25\mathrm{Var}(Z)\\ &\quad -2\cdot 2\mathrm{Cov}(X,Y)+2\cdot 5\mathrm{Cov}(X,Z)-2\cdot 10\mathrm{Cov}(Y,Z)\\ &= 892. \quad \Box\end{aligned}$$

8. Suppose $X \sim \mathrm{Exp}(\lambda)$. Use the m.g.f. of X to find $\mathrm{E}[X^k]$.

   Solution: By class notes, the m.g.f. of the $\mathrm{Exp}(\lambda)$ is $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda > t$. Therefore,

$$\mathrm{E}[X^k] = \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0} = \frac{k!}{\lambda^k},$$

   where the final answer follows after a little elbow grease. $\Box$.

9. (Hines et al., 4–18.) Let X and Y be two random variables such that $Y = a + bX$. Show that the moment generating function of Y is $M_Y(t) = e^{at} M_X(bt)$.

   Solution:

$$M_Y(t) = \mathrm{E}[e^{tY}] = \mathrm{E}[e^{t(a+bX)}] = e^{at}\mathrm{E}[e^{(bt)X}] = e^{at} M_X(bt). \quad \Box$$

10. (Hines et al., 5–2, binomial.) Six independent trips to the moon are planned, each of which has estimated success probability 0.95. What's the probability that at most 4 will be successful?

    Solution:

$$\begin{aligned}P(X \le 5) &= 1 - \sum_{x=5}^{6} \binom{6}{x} (0.95)^x (0.05)^{6-x}\\ &= 1 - 6(0.95)^5(0.05) + (0.95)^6\\ &= 1 - 0.9672 = 0.0328. \quad \Box\end{aligned}$$

11. (Hines et al., 5–6, binomial m.g.f.) Find the mean and variance of the binomial using the moment generating function.

Solution:

$$
\begin{aligned}
M_X(t) &= \mathrm{E}[e^{tX}] = \sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x(1-p)^{n-x} = (pe^t+q)^n, \qquad \text{where } q = 1-p. \\
\mathrm{E}[X] &= M_X^{'}(0) = [n(pe^t+q)^{n-1}pe^t]|_{t=0} = np. \\
\mathrm{E}[X^2] &= M_X^{''}(0) = np[e^t(n-1)(pe^t+q)^{n-2}(pe^t)+(pe^t+q)^{n-1}e^t]|_{t=0} \\
&= (np)^2 - np^2 + np. \\
\mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq. \quad \Box
\end{aligned}
$$

12. (Hines et al., 5–9, geometric.) The probability of a successful firing of a cruise missile is 0.95. Assuming independent tests, what's the prob that the first failure occurs with the fifth missile?

    Solution: $P(X=5) = (0.95)^4(0.05) = 0.0407. \quad \Box$

13. (Hines et al., 5–30, Poisson.) Phone calls arrive at a switchboard according to a Pois(10/hour) process. The current system can handle up to 20 calls in an hour without becoming overloaded. What's the probability of an overload in the next hour?

    Solution:

$$
\begin{aligned}
P(X>20) = P(X\geq 21) &= \sum_{x=21}^{\infty} \frac{e^{-10}(10)^x}{x!} \\
&= 1 - P(X\leq 20) = 1 - \sum_{x=0}^{20} \frac{e^{-10}(10)^x}{x!} \\
&= 0.002. \quad \Box
\end{aligned}
$$

14. (Hines et al., 6–13, exponential.) The time to failure of a TV is exponential with a mean of 3 years. A company offers insurance for the first year of usage. On what percentage of policies will the company have to pay claims?

    Solution: Let X = Life Length.

$$
\mathrm{E}(X) = \frac{1}{\lambda} = 3 \quad \Rightarrow \quad \lambda = \frac{1}{3},
$$

so

$$P(X < 1) \;=\; 1 - e^{-1/3} \;=\; 0.283.$$

Thus, 28.3% of policies result in a claim. $\Box$

15. (Hines et al., 6–16, exponential.) A transistor has an exponential time-to-failure distribution with a mean-time-to-failure of 20,000 hours. Suppose that the transistor has already lasted 20,000 hours. What's the probability that it fails by 30,000 hours?

Solution:

$$P(X > x+s|X > x) \;=\; P(X > s) \;=\; P(X > 10000) \;=\; e^{-10000/20000} \;=\; 0.6064,$$

so $P(X < 30000|X > 20000) \;=\; 0.3936.$ $\Box$

16. (Hines et al., 7–1(a)–(e), normal.) Suppose Z is standard normal. Find

    (a) $P(0 < Z < 2)$.
    (b) $P(-1 < Z < 1)$.
    (c) $P(Z < 1.65)$.
    (d) $P(Z > -1.96)$.
    (e) $P(|Z| > 1.5)$.

Solution:

(a) $P(0 \le Z \le 2) = \Phi(2) - \Phi(0) = 0.97725 - 0.5 = 0.47725$.
(b) $P(-1 \le Z \le 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.68268$.
(c) $P(Z \le 1.65) = \Phi(1.65) = 0.95053$.
(d) $P(Z \ge -1.96) = \Phi(1.96) = 0.9750$.
(e) $P(|Z| \ge 1.5) = 2[1 - \Phi(1.5)] = 0.1336$. $\Box$

17. (Hines et al., 7–3(a), normal.) Find c such that $\Phi(c) = 0.94062$.

Solution: From the back of the book, $c = \Phi^{-1}(0.94062) = 1.56$. $\Box$

18. (Hines et al., 7–5(a), normal.) If $X \sim N(80, 10^2)$, find $P(X < 100)$.

Solution: $P(X \le 100) = \Phi\left(\frac{100 - 80}{10}\right) = \Phi(2) = 0.97725$. $\Box$

19. (Hines et al., 7–7, normal.) A manager requires job applicants to take a test and score a 500. The test scores are normally distributed with a mean of 485 and standard deviation of 30. What percent of applicants pass?

Solution: $P(X > 500) = 1 - \Phi\left(\frac{500-485}{30}\right) = 1 - \Phi(0.5) = 0.30854$, i.e., 30.854%. $\Box$

20. **Mathemusical Bonus:** Suppose that a, k, and e are all nonzero. Use Beatles lyrics to prove that $m = t$.

Solution: According to "The End", the Beatles state that "The love you take is equal to the love you make." Canceling all of the similar terms and dividing by ake, we obtain the desired result. $\Box$

21. (Hines et al., 7-26. CLT.) 100 small bolts are packed in a box. Each weighs an average of 1 ounce, with a standard deviation of 0.1 ounce. Find the probability that a box weighs more than 102 ounces.

Solution: Let X_i be the weight of the ith bolt and let $Y = \sum_{i=1}^{100} X_i$ be the weight of the box. Note that $\mathrm{E}(X_i) = 1$, $\mathrm{Var}(X_i) = 0.01$, $i = 1, 2, \ldots, 100$.

Assuming that the X_i's are independent, we use the central limit theorem to approximate the distribution of $Y \sim \mathrm{Nor}(100, 1)$. Then

$$P(Y > 102) = P\left(Z > \frac{102 - 100}{1}\right) = 1 - \Phi(2) = 0.02275. \quad \Box$$

22. (Hines et al., 7–29(a). CLT.) A production process produces items, of which 8% are defective. A random sample of 200 items is selected every day and the number of defective items X is counted. Using the normal approximation to the binomial, find $P(X \le 16)$.

Solution: $p = 0.08$, $n = 200$, $np = 16$, $\sqrt{npq} = 3.84$. Let's incorporate the "continuity correction," and then the CLT:

$$\begin{aligned} P(X \le 16) &= P(X \le 16.5) \\ &\approx P\left(Z \le \frac{16.5 - np}{\sqrt{npq}}\right) \quad (\text{where } Z \sim \mathrm{Nor}(0,1)) \\ &= P\left(Z \le \frac{16.5 - 16}{3.84}\right) \\ &= \Phi(0.13) = 0.55172. \quad \Box \end{aligned}$$

23. (Hines et al., 7–37. lognormal.) The random variable $Y = \ell n(X)$ has a $\text{Nor}(50, 25)$ distribution. Find the mean, variance, mode, and median of X.

Solution: I got these answers by directly plugging into the equations from the book. For example, in general, $\text{E}[X] = \exp(\mu + \sigma^2/2) = e^{62.5}$. And similarly, $\text{Var}(X) = e^{125}(e^{25} - 1)$, $\text{median}(X) = e^{50}$, $\text{mode}(X) = e^{25}$. □

24. Computer Exercises — Random Variate Generation

(a) Let's start out with something easy — the Uniform(0,1) distribution. To generate a Uniform(0,1) random variable in Excel, you simply type `= RAND()`. Copy an entire column of 100 of these guys and make a histogram. If things don't look particularly uniform, try the same exercise for 1000 observations. By the way, you can use the `<F9>` key to get an independent realization of your experiment.

(b) It's very easy to generate an Exponential(1) random variable in Excel. Just use

$$= -\texttt{LN}(\texttt{RAND}())$$

(This result uses the inverse transform method from Module 2.6.) Generate 1000 or so of these guys and make a nice histogram.

(c) In Excel, you can generate a Normal(0,1) random variable using

$$= \texttt{NORMINV}(\texttt{RAND}(), \texttt{0}, \texttt{1}) \qquad \text{(inverse transform method)}$$

or

$$= \texttt{SQRT}(-\texttt{2} * \texttt{LN}(\texttt{RAND}())) * \texttt{COS}(\texttt{2} * \texttt{PI}() * \texttt{RAND}()) \qquad \text{(Box–Muller method)}$$

Generate a bunch of normals using one of the above equations and make a histogram.

(d) Triangular distribution. Generate two columns of Uniform(0,1)'s. In the third column, add up the respective entries from the previous two columns, e.g., `C1 = A1 + B1`, etc. Make a histogram of the third column. Guess what you get?

Solution: You get a triangular p.d.f. Surprise! □

(e) Normal distribution from the Central Limit Theorem. Generate *twelve* columns of Uniform(0,1)'s. In the 13th column, add up the respective entries from the previous 12 columns. Make a histogram of the 13th column. Guess what you get this time?

Solution: You get what looks like a normal p.d.f. The CLT works! □

(f) Cauchy distribution. It turns out that you can generate a Cauchy random variable as the ratio of two i.i.d. Nor(0,1)'s. Make a histogram and comment. Does the CLT work for this distribution?

Solution: You get a mess that has extreme values. If you zoom in towards $x = 0$, it looks vaguely normal — but the tails are way too fat to actually be normal. If you try to apply the CLT, it fails — in fact, you get another Cauchy. The reason for the CLT failure is that the variance of the Cauchy is infinite, thus violating one of the CLT assumptions. □

Most of the following problems are from Hines, et al.

10–40(a). The life in hours of a 75-W light bulb is known to be approximately normally distributed, with a standard deviation of $\sigma = 25$ hours. A random sample of 20 bulbs has a mean life of $\bar{x} = 1014$ hours. Construct a 95% two-sided confidence interval on the mean life.

Solution: Since σ is known, we use

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Since $z_{0.025} = 1.96$, we have $1003.04 \le \mu \le 1024.96$. □

10–42. Suppose that in Exercise 10–40 we wanted to be 95% confident that the error in estimating the mean life is less than 5 hours. What sample size should be used?

Solution: $n = (z_{\alpha/2}\sigma/\epsilon)^2 = [(1.96)25/5]^2 = 96.04 \simeq 97.$ □

10–46. The burning rates of two different solid-fuel rocket propellants are being studied. It is known that both propellants have approximately the same standard deviation of burning rate, $\sigma_1 = \sigma_2 = 3$ cm/s. Two random samples of $n_1 = 20$ and $n_2 = 20$ specimens are tested, and the sample mean burning rates are $\bar{x}_1 = 18$ and $\bar{x}_2 = 24$ cm/s. Construct a 99% confidence interval on the mean difference in burning rate.

Solution: Since both variances are known, we use

$$\bar{x}_2 - \bar{x}_1 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_2 - \bar{x}_1 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Since $z_{0.005} = 2.576$, we have $3.56 \le \mu_2 - \mu_1 \le 8.44$. □

10–48(a). The compressive strength of concrete is being tested by a civil engineer. He tests 16 specimens and obtains the following data:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295
2250	2238	2300	2217

Construct a 95% two-sided confidence interval on the mean strength.

Solution: Since σ is unknown, we use

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}.$$

We can easily calculate

$$\bar{x} = 2257.75 \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = (34.51)^2.$$

Since $t_{0.025,15} = 2.13$, we have $2239.4 \le \mu \le 2276.1$. □

10–49. An article in *Annual Reviews Material Research* (2001, p. 291) presents bond strengths for various energetic materials (explosives, propellants, and pyrotechnics). Bond strengths for 15 such materials are shown below. Construct a two-sided 95% confidence interval on the mean bond strength.

323	312	300	284	283
261	207	183	180	179
174	167	167	157	120

Solution: Since σ is unknown, we use

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}.$$

We can easily calculate

$$\bar{x} = 219.80 \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = (66.41)^2.$$

Since $t_{0.025,14} = 2.14$, we have $183.1 \le \mu \le 256.5$. $\Box$

10–50. The wall thickness of 25 glass 2-liter bottles was measured by a quality-control engineer. The sample mean was $\bar{x} = 4.05$ mm, and the sample standard deviation was $s = 0.08$ mm. Find a 90% lower confidence interval on the mean wall thickness.

Solution: The confidence interval will have the form

$$\bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} \;\le\; \mu$$

Since $t_{0.10,24} = 1.32$, we have $4.05 - t_{0.10,24}(0.08/\sqrt{25}) \le \mu$. In other words, $4.029 \le \mu$. $\Box$

10–56(a). Random samples of size 20 were drawn from two independent normal populations. The sample means and standard deviations were $\bar{x}_1 = 22.0$, $s_1 = 1.8$, $\bar{x}_2 = 21.5$, and $s_2 = 1.5$. Assuming that $\sigma_1^2 = \sigma_2^2$, find a 95% two-sided confidence interval on $\mu_1 - \mu_2$.

Solution: Since both variances are *unknown but assumed equal*, we use

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},$$

where $n_1 = n_2 = 20$ and the pooled variance is

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 2.745.$$

Since $t_{0.025,38} = 2.024$, we have $-0.561 \le \mu_1 - \mu_2 \le 1.561$. $\Box$

10–57. The diameter of steel rods manufactured on two different extrusion machines is being investigated. Two random samples of sizes $n_1 = 15$ and $n_2 = 18$ are selected, and the sample means and sample variances are $\bar{x}_1 = 8.73$, $s_1^2 = 0.30$, $\bar{x}_2 = 8.68$, and $s_2^2 = 0.34$. Assuming that $\sigma_1^2 = \sigma_2^2$, construct a 95% two-sided confidence interval on the difference in mean rod diameter.

Solution: Using the same equations as in the solution to Question 10–56(a), we obtain $-0.355 \le \mu_1 - \mu_2 \le 0.455$. (Note that the answer in the back of the book was wrong.) $\Box$

10–59(a). Consider the data in Exercise 10–48. Construct a 95% two-sided confidence interval on σ^2.

Solution: The desired confidence interval is of the form

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}.$$

From the solution to Exercise 10–48, we know that $s^2 = (34.51)^2$. Further, $\chi^2_{0.975,15} = 6.26$ and $\chi^2_{0.025,15} = 27.49$. Thus, the c.i. is $649.84 \le \sigma^2 \le 2853.69$. □

10–63. Consider the data in Exercise 10–56. Construct a 95% two-sided confidence interval on the ratio of the population variances σ_1^2/σ_2^2.

Solution: The desired confidence interval is of the form

$$\frac{s_1^2}{s_2^2}\,\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} \;\le\; \frac{\sigma_1^2}{\sigma_2^2} \;\le\; \frac{s_1^2}{s_2^2}\,F_{\alpha/2,n_2-1,n_1-1}.$$

In other words, we want

$$\frac{(1.8)^2}{(1.5)^2}\,\frac{1}{F_{0.025,19,19}} \;\le\; \frac{\sigma_1^2}{\sigma_2^2} \;\le\; \frac{(1.8)^2}{(1.5)^2}\,F_{0.025,19,19}.$$

Since $F_{0.025,19,19} = 2.526$, we obtain the c.i. $0.57 \le \sigma_1^2/\sigma_2^2 \le 3.64$. □

Bernoulli Question. A pollster asked a sample of 2000 people whether or not they were in favor of a particular proposal. Exactly 1200 people answered yes. Find a 95% confidence interval for the percentage of the population in favor of the proposal.

Solution: We are looking for a c.i. for the proportion p of favorable responses, i.e., the Bernnoulli parameter. Thus, the solution is of the form

$$\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \;\le\; p \;\le\; \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}.$$

That is,

$$0.6 - 1.96\sqrt{\frac{0.6(0.4)}{2000}} \;\le\; p \;\le\; 0.6 + 1.96\sqrt{\frac{0.6(0.4)}{2000}},$$

or $0.579 \le p \le 0.621$. □

BONUS: What do Stiller and Meara, Lou Reed, Suzanne Pleshette, and 44 have in common?

Solution: Syracuse University. GO ORANGE! □