Essential Mathematics for Political and Social Research

ANSWER KEY

Version 2.0 (March 5, 2007)†

Jeff Gill

Cambridge University Press

Cambridge, New York

† Thanks in particular to Cindy Kam, Tim McDaniel, and Ellie Powell for finding errors and typos in an earlier edition of this answer key.

1

The Basics

1.1 Simplify the following expressions as much as possible:

$$(-x^4y^2)^2 = x^8y^4 \qquad 9(3^0) = 9 \qquad (2a^2)(4a^4) = 8a^6$$

$$\frac{x^4}{x^3} = x \qquad (-2)^{7-4} = -8 \qquad \left(\frac{1}{27b^3}\right)^{1/3} = \frac{1}{3b}$$

$$y^7y^6y^5y^4 = y^{22} \qquad \frac{2a/7b}{11b/5a} = \frac{10a^2}{77b^2} \qquad (z^2)^4 = z^8$$

1.3 Solve:

$$\sqrt[3]{2^3} = 2 \qquad \sqrt[3]{27} = 3 \qquad \sqrt[4]{625} = 5$$

1.5 Another way to describe a line in Cartesian terms is the **point-slope form**: $(y - y') = m(x - x')$, where y' and x' are given values and m is the slope of the line. Show that this is equivalent to the form given by solving for the intercept.

Start with a basic algebraic rearrangement:

$$\begin{aligned}(y - y') &= m(x - x')\\ y - y' &= mx - mx'\\ y &= mx - mx' + y'\\ y &= mx + (-mx' + y').\end{aligned}$$

Since we know $y = mx + b$ where b is the y-intercept and m is the slope of the line then:

$$\begin{aligned}mx + b &= mx + (-mx' + y')\\ b &= (-mx' + y')\\ y' &= mx' + b.\end{aligned}$$

1.7 A very famous sequence of numbers is called the Fibonacci sequence, which starts with 0 and 1 and continues according to:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

Figure out the logic behind the sequence and write it as a function using subscripted values like x_j for the jth value in the sequence.

We want to explain the series: $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$, and it is easy to see that numbers are produced from adding consecutive values. This is described by:

$$x_n = x_{n-1} + x_{n-2},$$

for all $n \geq 2$ with the following conditions:

$$x_0 = 0, \qquad x_1 = 1$$

1.9 Which of the following functions are continuous? If not, where are the discontinuities?

$$f(x) = \frac{9x^3 - x}{(x-1)(x+1)} \qquad g(y,z) = \frac{6y^4z^3 + 3y^2z - 56}{12y^5 - 3zy + 18z}$$

$$f(x) = e^{-x^2} \qquad f(y) = y^3 - y^2 + 1$$

$$h(x,y) = \frac{xy}{x+y} \qquad f(x) = \begin{cases} x^3 + 1 \; x > 0 \\ \frac{1}{2} \; x = 0 \\ -x^2 x < 0 \end{cases}$$

(i) discontinuous at x=1,x=-1

(ii) all ratios of polynomials are continuous unless a denominator is zero ($y = 0, z = 0$ here).

(iii) continuous

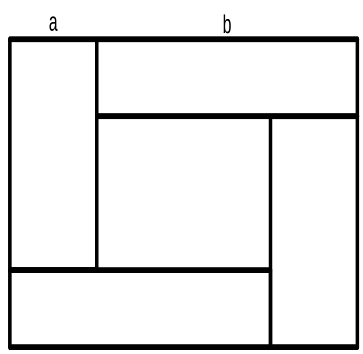
(iv) all polynomials are continuous

(v) discontinuous at (0,0);

(vi) discontinuous at x=0

1.11 Use the diagram of the square to prove that $(a-b)^2 + 4ab = (a+b)^2$

(i.e., demonstrate this equality geometrically rather than algebraically with features of the square shown).



(i) The right-hand side, $(a+b)^2$ gives the area of the full square since each side is of length $a+b$.

(ii) Notice that there are four equal sized rectangles and a smaller square that fill the entire full (large) square.

(iii) The rectangles are of size ab, thus their total contribution is $4ab$.

(iv) The smaller square has sides of size $b - a$, this is of size $(b - a)^2$.

(v) Therefore we know that $(a - b)^2 + 4ab = (a + b)^2$ (the square on the first term makes it irrelevant which value gets subtracted from the other).

1.13 Sørensen's (1977) model of social mobility looks at the process of increasing attainment in the labor market as a function of time, personal qualities, and opportunities. Typical professional career paths follow a logarithmic-like curve with rapid initial advancement and tapering off progress later. Label y_t the attainment level at time period t and y_{t-1} the attainment in the previous period, both of which are defined over $\mathfrak{R}^+$. Sørensen stipulates:

$$y_t = \frac{r}{s}[\exp(st) - 1] + y_{t-1}\exp(st),$$

where $r \in \mathfrak{R}^+$ is the individual's resources and abilities and $s \in \mathfrak{R}^+$ is the structural impact (i.e., a measure of opportunities that become available). What is the domain of s, that is, what restrictions are necessary on what values it can take on in order for this model to make sense in that declining marginal manner?

The model does not make sense unless $-r < s < 0$ since there is a theoretical requirement that effects are declining in increments over time. To see this, rearrange the formula to get $y_t = \exp(st)(\frac{r}{s} + y_{t-1}) - \frac{r}{s}$. This makes it clear that $|r| < |t|$ is necessary, and since t is a set of integers increasing by one, then a negative value of s is required to get diminishing positive values from the $\exp()$ function.

Some illustrative `R` code:

```
r <- 10; s <- -1/2; y <- 1
for (i in 2: 10) {
    y <- c( y, (r/s)*(exp(s*i)-1) + y[1]*exp(s*i) )
}
plot(y,type="l")
```

1.15 Using the change of base formula for logarithms, change $\log_6(36)$ to $\log_3(36)$.

$$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$$

$$\log_a(x) = \frac{\log_6(36)}{\log_6(3)} = \log_3(36)$$

1.17 Sociologists Holland and Leinhardt (1970) developed measures for models of structure in interpersonal relations using ranked clusters. This approach requires extensive use of factorials to express personal choices. The authors defined the notation $x^{(k)} = x(x-1)(x-2)\cdots(x-k+1)$. Show that $x^{(k)}$ is just $x!/(x-k)!$.

$$x! = x(x-1)(x-2)\cdots(x-k+1)(x-k)\cdots 3\cdot 2\cdot 1$$

$$\frac{x!}{(x-k)!} = \frac{x(x-1)(x-2)\cdots(x-k+1)(x-k)\cdots 3\cdot 2\cdot 1}{(x-k)!}$$

$$= \frac{x(x-1)(x-2)\cdots(x-k+1)\,[(x-k)\cdots 3\cdot 2\cdot 1]}{(x-k)!}$$

$$= \frac{x(x-1)(x-2)\cdots(x-k+1)\,[(x-k)!]}{(x-k)!}$$

$$= x(x-1)(x-2)\cdots(x-k+1)$$

$$= x^{(k)}$$

1.19 Show that in general

$$\sum_{i=1}^{m}\prod_{j=1}^{n} x_i y_j \neq \prod_{j=1}^{n}\sum_{i=1}^{m} x_i y_j$$

and construct a special case where it is actually equal.

$$\sum_{i=1}^{m}\prod_{j=1}^{n} x_i y_j = [x_1(y_1)(y_2),\ldots,(y_n)] + [x_2(y_1)(y_2),\ldots,(y_n)] +$$

$$\ldots + [x_m(y_1)(y_2),\ldots,(y_n)]$$

$$\prod_{j=1}^{n}\sum_{i=1}^{m} x_i y_j = [(y_1)(x_1 + \ldots + x_m)] \cdot [(y_2)(x_1 +$$

$$\ldots + x_m)] \cdot \ldots \cdot [(y_n)(x_1 + \ldots + x_m)]$$

$$= (x_1 + \ldots + x_m)^n [(y_1)(y_2),\ldots,(y_n)]$$

So: $(x_1 + \ldots + x_m)[(y_1)(y_2),\ldots,(y_n)]$

$$\neq (x_1 + \ldots + x_m)^n [(y_1)(y_2),\ldots,(y_n)]$$

1.21 Suppose we had a linear regression line relating the size of state-level unemployment percent on the x-axis and homicides per 100,000 of the state population on the y-axis, with slope $m = 2.41$ and intercept $b = 27$. What would be the expected effect of increasing unemployment by 5%?

For 5-unit (percent) positive change in state-level unemployment, we get the following expected change in homicides per 100,000:

$$\delta Y = m(\delta X) + b$$
$$= 2.41(5) + 27$$
$$= 39.05$$

1.23 Use Euler's expansion to calculate e with 10 terms. Compare this result to some definition of e that you find in a mathematics text. How accurate were you?

$$1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} + \frac{1}{7!} + \frac{1}{8!} + \frac{1}{9!} + \frac{1}{10!}$$
$$+ \frac{1}{11!} + \frac{1}{12!} + \frac{1}{13!} + \frac{1}{14!} + \frac{1}{15!} + \frac{1}{16!} + \frac{1}{17!}$$
$$+ \frac{1}{18!} + \frac{1}{19!} + \frac{1}{20!}$$
$$= \frac{6613313319248080001}{2432902008176640000} = 2.718281828$$

1.25 Find the roots (solutions) to the following quadratic equations:

$$4x^2 - 1 = 17$$

$$4x^2 = 18, x^2 = 4 \quad \Rightarrow \; x = \pm\sqrt{9/2}$$

$$9x^2 - 3x - 12 = 0$$

$$(3x+3)(3x-4) = 0 \quad \Rightarrow \; x = -1, \frac{4}{3}$$

$$x^2 - 2x - 16 = 0$$

$$\frac{1}{2a}\left(-b \pm \sqrt{(b^2 - 4ac)}\right) = \frac{1}{2}\left(2 \pm \sqrt{(2^2 - (4 * (-16))}\right)$$
$$= \frac{1}{2}\left(2 \pm \sqrt{4 - (-64)}\right) \qquad \Rightarrow \; x = \pm\sqrt{17} + 1$$

$$6x^2 - 6x - 6 = 0$$

$$\frac{1}{12}\left(-6 \pm \sqrt{(6^2 - (4 * (-36))}\right) = \frac{1}{12}\left(-6 \pm \sqrt{36 - (-144)}\right)$$
$$\Rightarrow \; \pm\frac{1}{2}\sqrt{5} - \frac{1}{2}$$

$$5 + 11x = -3x^2$$

$$3x^2 + 11x + 5 = 0$$

$$\frac{1}{6}\left(-11 \pm \sqrt{(11^2 - (4 * 3 * 5)}\right) \qquad \Rightarrow \; \pm\frac{1}{6}\sqrt{61} - \frac{11}{6}$$

1.27 The Nachmias–Rosenbloom Measure of Variation (MV) indicates how many heterogeneous intergroup relationships are evident from the full set of those mathematically possible given the population. Specifically it is described in terms of the "frequency" (their original language) of observed subgroups in the full group of interest. Call f_i the frequency or proportion of the ith subgroup and n the number of these groups. The index is created by

$$MV = \frac{\text{“each frequency} \times \text{all others, summed”}}{\text{“number of combinations”} \times \text{“mean frequency squared”}}$$

$$= \frac{\sum_{i=1}^{n}(f_i \neq f_j)}{\frac{n(n-1)}{2}\bar{f}^2}.$$

Nachmias and Rosenbloom (1973) use this measure to make claims about how integrated U.S. federal agencies are with regard to race. For a population of 24 individuals:

a) What mixture of two groups (say blacks and whites) gives the maximum possible MV? Calculate this value.

Obviously this number is maximized when there are an equal number of people in each subgroup, and minimized when one group is as large as possible and all others have just one person. For example, consider a population of 24 in which 12 are black and 12 are white. For this organization the Measure of Variation would be calculated by:

$$MV = \frac{12 \times 12}{\frac{2}{2} \times 12^2} = 1.$$

b) What mixture of two groups (say blacks and whites) gives the minimum possible MV but still has both groups represented? Calculate this value as well.

Conversely, the most disproportionately possible organization would have only one black. This organization would have a Measure of Variation:

$$MV = \frac{23 \times 1}{\frac{2}{2} \times 12^2} = 0.16.$$

2

Analytic Geometry

2.1 For the following values of θ in radians or degrees (indicated), calculate $\sin(\theta)$, $\cos(\theta)$, and $\tan(\theta)$:

0	$\frac{3\pi}{2}$	$\frac{11\pi}{4}$	80π
$\frac{22}{7}$	150^o	$-\frac{2\pi}{5}$	390^o
$80,000\pi$	-4π	$3\pi + .001$	10^o

$$\sin(0) = 0,\ \cos(0) = 1,\ \tan(0) = 0$$

$$\sin\left(\frac{3\pi}{2}\right) = 1,\ \cos\left(\frac{3\pi}{2}\right) = 0,\ \sin\left(\frac{3\pi}{2}\right) = \infty$$

$$\sin\left(\frac{11\pi}{4}\right) = 0.7071068,\ \cos\left(\frac{11\pi}{4}\right) = -0.7071068,$$
$$\tan\left(\frac{11\pi}{4}\right) = -1$$

$$\sin(80\pi) = 0,\ \cos(80\pi) = 1,\ \tan(80\pi) = 0$$

$$\sin\left(\frac{22}{7}\right) = -0.001264489,\ \cos\left(\frac{22}{7}\right) = -0.999999201,$$
$$\tan\left(\frac{22}{7}\right) = 0.001264490$$

$$\sin(150^o) = 0.5,\ \sin(150^o) = -0.8660254,$$
$$\sin(150^o) = -0.5773503$$

$$\sin\left(-\frac{2\pi}{5}\right) = 0.9510565,\ \sin\left(-\frac{2\pi}{5}\right) = 0.3090170,$$
$$\sin\left(-\frac{2\pi}{5}\right) = 3.0776835$$

$$\sin(390^o) = 0.5,\ \sin(390^o) = 0.8660254,\ \sin(390^o) = 0.5773503$$

$$\sin(80,000\pi) = 0,\ \sin(80,000\pi) = 1,\ \sin(80,000\pi) = 0$$

$$\sin(-4\pi) = 0,\ \sin(-4\pi) = 1,\ \sin(-4\pi) = 0$$

$$\sin(3\pi + .001) = -0.0009999998,\ \sin(3\pi + .001) = -0.9999995,$$

$$\sin(3\pi + .001) = 0.0010000003$$

$$\sin(10^o) = 0.1736482,\ \sin(10^o) = 0.9848078,$$
$$\sin(10^o) = 0.1763270$$

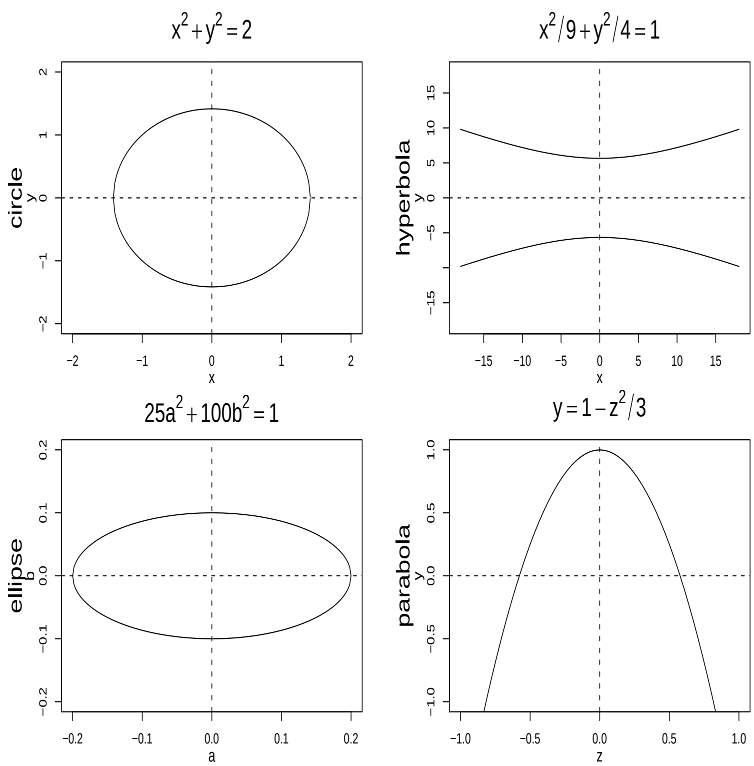
2.3 Show that $\cos(\theta)\tan(\theta) = \sin(\theta)$.

Since $\tan(\theta) = \frac{\text{opposite}}{\text{adjacent}}$, it equals $\frac{\text{opposite}}{\text{hypotenuse}} / \frac{\text{hypotenuse}}{\text{opposite}}$. This is $\sin(\theta)/\cos(\theta)$ so $\cos(\theta) \times \sin(\theta)/\cos(\theta) = \sin(\theta)$.

2.5 Identify and sketch the conic sections:

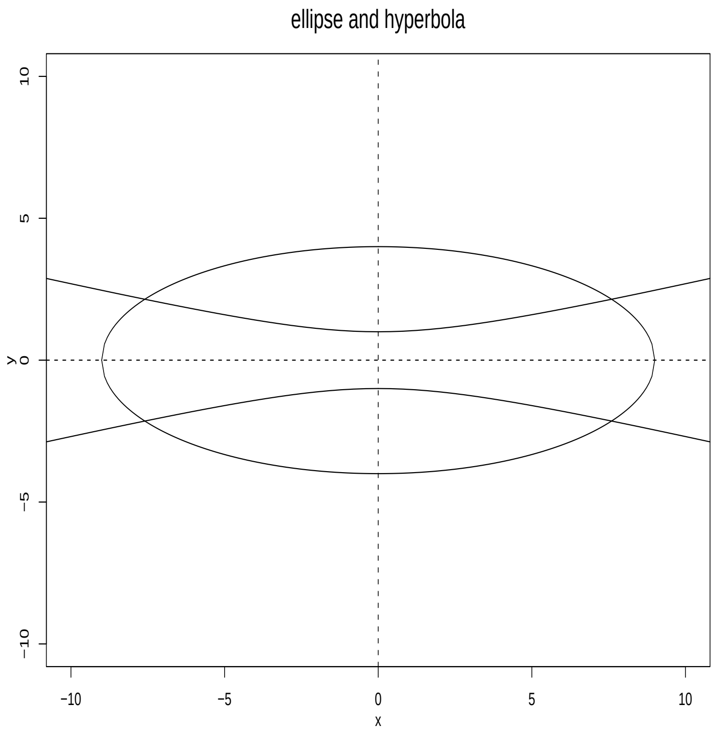
$$x^2 + y^2 = 2 \qquad \frac{x^2}{9} - \frac{y^2}{4} = 1$$

$$25a^2 + 100b^2 = 1 \qquad y = 1 - \frac{z^2}{3}$$



2.7 Draw in the same figure an ellipse with $a = 9$, $b = 4$, and a hyperbola with $a = 4$, $b = 1$. Calculate the four points of intersection.

Restricting x and y to be positive, the two equations can be solved for two



unknowns producing $x = 7.595094$ and $y = 2.146006$. By symmetry the other points of intersection are $[-7.595094, 2.146006]$, $[7.595094, -2.146006]$, $[-7.595094, -2.146006]$.

2.9 Data analysts sometimes find it useful to transform variables into more convenient forms. For instance, data that are bounded by $[0\!:\!1]$ can sometimes be inconvenient to model due to the bounds. Some transformations that address this (and can be easily undone to revert back to the original form) are

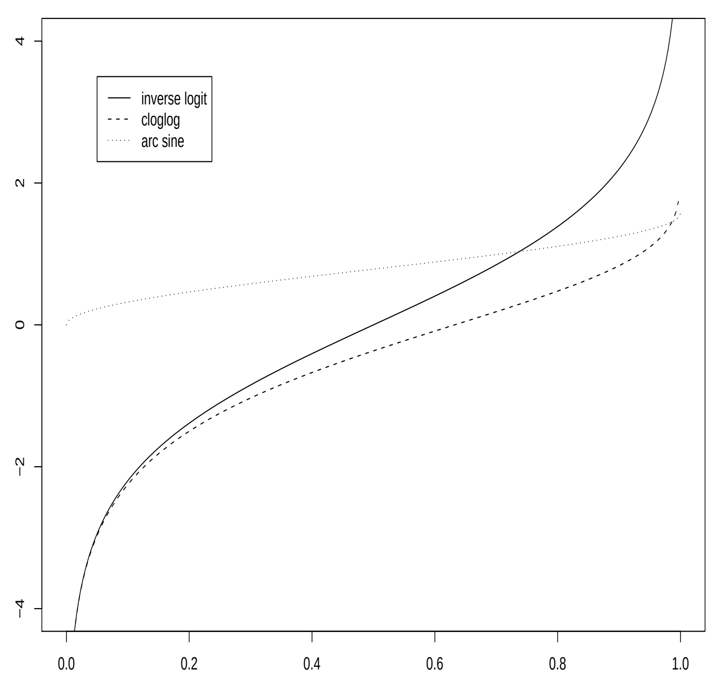
$$\text{inverse logit: } f(x) = \log(x/(1-x))$$

$$\text{cloglog: } f(x) = \log(-\log(1-x))$$

$$\text{arc sine: } f(x) = \arcsin(\sqrt{x}).$$

For $x \in [0:1]$, what advantage does the arcsine transformation have? Graph each of these functions over this domain in the same graph. If you were modeling the underlying preference or utility for a dichotomous choice (0 and 1, such as a purchase or a vote), which form would you prefer? Write the inverse function to the arcsine transformation. Note: The arcsine transformation has been useful in analysis of variance (ANOVA) statistical tests when distributional assumptions are violated.

From the figure it is clear that the inverse logit transformation has the most curvature and therefore may better describe preference changes. It also has defined endpoints which the cloglog does not. The arc sine transformation shows little curvature except at the endpoints. The inverse of the arc sine transformation is $f(y) = (\sin(y))^2$.



2.11 Suppose you had an ellipse enclosed within a circle according to:

$$\text{ellipse:}\quad \frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \qquad a>b$$
$$\text{circle:}\quad x^2+y^2=a^2.$$

What is the area of the ellipse?

$$ab\pi = area, a^2 = x^2+y^2, \frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$
$$a=\sqrt{x^2+y^2}$$
$$\frac{x^2}{(\sqrt{x^2+y^2})^2}+\frac{y^2}{b^2}=1$$
$$\frac{x^2}{x^2+y^2}+\frac{y^2}{b^2}=1$$
$$\frac{y^2}{b^2}=1-\frac{x^2}{x^2+y^2}$$
$$\frac{y^2}{b^2}=\frac{y^2}{x^2+y^2}$$
$$\frac{1}{b^2}=\frac{1}{x^2+y^2}$$
$$b=\sqrt{x^2+y^2}$$
$$\Rightarrow: ab\pi = \left(x^2+y^2\right)\pi$$

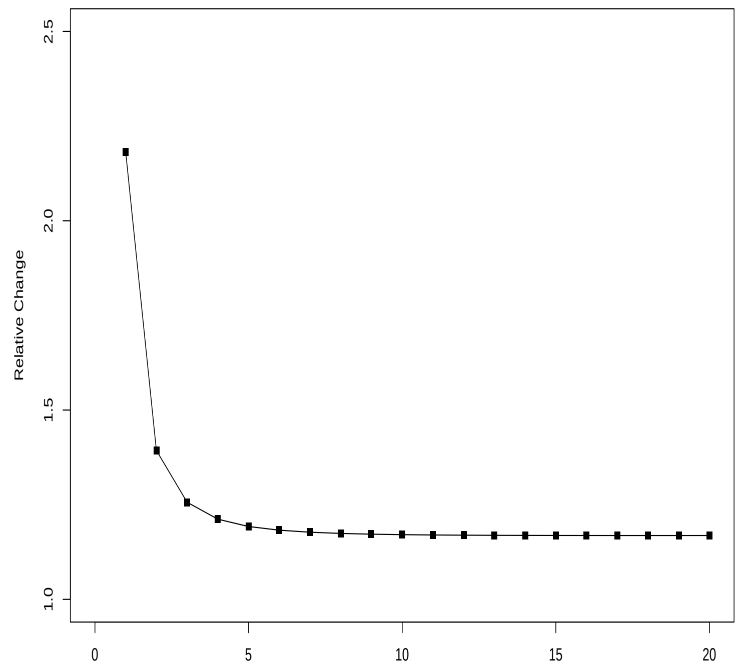
2.13 Yeats, Irons, and Rhoades (1975) found that annual deposit growth for 48 commercial banks can be modeled by the function:

$$\frac{D_{t+1}}{D_t}=1.172-0.125t^{-1}+1.135t^{-2},$$

where D is year-end deposits and t is years. Graph this equation for 20 years and identify the form of the curve.

This is a section of the hyperbolic form. See graph below.

2.15 In studying the labor supply of nurses, Manchester (1976) defined ν as the wage and $H(\nu)$ as the units of work per time period (omitting a constant term, which is unimportant to the argument). He gave two possible explanatory functions: $H(\nu) = a\nu^2 + b\nu$ with constants $a<0, b>0$, and $H(\nu) =$



$a+b/\nu$ with constants a unconstrained, $b < 0$. Which of these is a hyperbolic function? What is the form of the other?

The second one is the hyperbolic form and the first is quadratic.

3

Linear Algebra: Vectors, Matrices, and Operations

3.1 Perform the following vector multiplication operations:

$$[1 \quad 1 \quad 1] \cdot [a \quad b \quad c]'$$

$$[1 \quad 1 \quad 1] \times [a \quad b \quad c]'$$

$$[-1 \quad 1 \quad -1] \cdot [4 \quad 3 \quad 12]'$$

$$[-1 \quad 1 \quad -1] \times [4 \quad 3 \quad 12]'$$

$$[0 \quad 9 \quad 0 \quad 11] \cdot [123.98211 \quad 6 \quad -6392.38743 \quad -5]'$$

$$[123.98211 \quad 6 \quad -6392.38743 \quad -5] \cdot [0 \quad 9 \quad 0 \quad 11]'.$$

- $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = (1)a + (1)b + (1)c = a + b + c$
- $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix} = [(1)(c) - (1)(b), (1)(c) - (1)(a), (1)(b) - (1)(a)] = [c - b, c - a, b - a]$

- $\begin{bmatrix} -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 12 \end{bmatrix} = (-1)(4) + (1)(3) + (-1)(12) = -13$
- $\begin{bmatrix} -1 & 1 & -1 \end{bmatrix} \times \begin{bmatrix} 4 \\ 3 \\ 12 \end{bmatrix} = [(1)(12)-(-1)(3), (-1)(12)-(-1)(4), (-1)(3)-(1)(4)] = [15, -8, -7]$
- $\begin{bmatrix} 0 & 9 & 0 & 11 \end{bmatrix} \begin{bmatrix} 123.98211 \\ 6 \\ -6392.38743 \\ -5 \end{bmatrix} = -1$
- $\begin{bmatrix} 123.98211 \\ 6 \\ -6392.38743 \\ -5 \end{bmatrix} \begin{bmatrix} 0 & 9 & 0 & 11 \end{bmatrix} = -1$

3.3 Show that $\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\|\mathbf{w}\|\cos\theta$ implies $\cos(\theta) = \frac{\mathbf{vw}}{\|\mathbf{v}\|\,\|\mathbf{w}\|}$.

$$
\begin{aligned}
\cos(\theta) &= \frac{\|\mathbf{v} - \mathbf{w}\|^2 - \|\mathbf{v}\|^2 - \|\mathbf{w}\|^2}{-2\|\mathbf{v}\|\|\mathbf{w}\|} \\
&= \frac{(\mathbf{v} - \mathbf{w})' \cdot (\mathbf{v} - we) - \mathbf{v}' \cdot \mathbf{v} - \mathbf{w}' \cdot \mathbf{w}}{-2\|\mathbf{v}\|\|\mathbf{w}\|} \\
&= \frac{\mathbf{v}' \cdot \mathbf{v} - \mathbf{v}' \cdot \mathbf{w} - \mathbf{w}'\mathbf{v} + \mathbf{w}'\mathbf{w} - \mathbf{v}'\mathbf{v} - \mathbf{w}'\mathbf{w}}{-2\|\mathbf{v}\|\|\mathbf{w}\|} \\
&= \frac{-2\mathbf{v}'\mathbf{w}}{-2\|\mathbf{v}\|\|\mathbf{w}\|}
\end{aligned}
$$

3.5 Explain why the perpendicularity property is a special case of the triangle inequality for vector p-norms.

For values of p ranging from 1 to ∞, perpendicular vectors give the range $[2k, k]$ for the triangle value $||\mathbf{u} + ve||_p$, where k is the normalized length of the vectors.

3.7 Show that pre-multiplication and post-multiplication with the identity matrix are equivalent.

Use a diagonal matrix to show a slightly more general property:

$$D = \begin{bmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nn} \end{bmatrix}$$

$$DX = \begin{bmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_n \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} d_1x_{11} & d_1x_{12} & ... & d_1x_{1n} \\ d_2x_{21} & d_2x_{22} & ... & d_2x_{2n} \\ ... & ... & ... & ... \\ d_nx_{n1} & d_nx_{n2} & ... & d_nx_{nn} \end{bmatrix}$$

$$XD = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nn} \end{bmatrix} \begin{bmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_n \end{bmatrix}$$

$$= \begin{bmatrix} d_1x_{11} & d_2x_{12} & ... & d_nx_{1n} \\ d_1x_{21} & d_2x_{22} & ... & d_nx_{2n} \\ ... & ... & ... & ... \\ d_1x_{n1} & d_2x_{n2} & ... & d_nx_{nn} \end{bmatrix}$$

If D is the identity matrix, then $d_j = 1$ for all $j = 1, 2, ..., n$ and,

$$\begin{bmatrix} d_1x_{11} & d_1x_{12} & ... & d_1x_{1n} \\ d_2x_{21} & d_2x_{22} & ... & d_2x_{2n} \\ ... & ... & ... & ... \\ d_nx_{n1} & d_nx_{n2} & ... & d_nx_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} d_1x_{11} & d_2x_{12} & ... & d_nx_{1n} \\ d_1x_{21} & d_2x_{22} & ... & d_nx_{2n} \\ ... & ... & ... & ... \\ d_1x_{n1} & d_2x_{n2} & ... & d_nx_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nn} \end{bmatrix}$$

3.9 For the following matrix, calculate $\mathbf{X}^n$ for $n = 2, 3, 4, 5$. Write a rule for calculating higher values of n.

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

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^3 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^4 = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 1 & 0 \\ 3 & 0 & 5 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^5 = \begin{bmatrix} 3 & 0 & 5 \\ 0 & 1 & 0 \\ 5 & 0 & 8 \end{bmatrix}$$

There are different rules that can be written, but the following covers a simple version:

$$\mathbf{X}^n[2,] < -\mathbf{X}^{n-1}[2,] \quad \text{and} \quad \mathbf{X}^n[,2] < -\mathbf{X}^{n-1}[,2]$$

$$\mathbf{X}^n[3,1] = \mathbf{X}^n[1,3] = \mathbf{X}^{n-1}[3,3]$$

$$\mathbf{X}^n[1,1] = \mathbf{X}^{n-1}[1,3]$$

$$\mathbf{X}^n[3,3] = \mathbf{X}^{n-1}[3,1] + X^{n-1}[3,3]$$

3.11 Perform the following matrix multiplications:

$$\begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 3 & 0 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 & -2 \\ 6 & 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 3 & 0 \\ 1 & 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -9 \\ -1 & -4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -4 & -4 \\ -1 & 0 \\ -3 & -8 \end{bmatrix}' \qquad \begin{bmatrix} 0 & 0 \\ 0 & \infty \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

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

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 3 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 5 & 9 \\ 7 & 7 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 & -2 \\ 6 & 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 3 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 13 & 17 \\ 37 & 50 \end{bmatrix}$$

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

$$\begin{bmatrix} -1 & -9 \\ -1 & -4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -4 & -4 \\ -1 & 0 \\ -3 & -8 \end{bmatrix}' = \begin{bmatrix} 40 & 1 & 75 \\ 20 & 1 & 35 \\ -12 & -1 & -19 \end{bmatrix}$$

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

3.13 Communication within work groups can sometimes be studied by looking analytically at individual decision processes. Roby and Lanzetta (1956) studied at this process by constructing three matrices: OR, which maps six observations to six possible responses; PO, which indicates which type of person from three is a source of information for each observation; and PR, which maps who is responsible of the three for each of the six responses. They give these matrices (by example) as

$$OR = \begin{array}{c} \\ O_1 \\ O_2 \\ O_3 \\ O_4 \\ O_5 \\ O_6 \end{array} \begin{array}{c} \begin{array}{cccccc} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

$$PO = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \end{array} \begin{array}{c} \begin{array}{cccccc} O_1 & O_2 & O_3 & O_4 & O_5 & O_6 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}.$$

$$PR = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \end{array} \begin{array}{c} \begin{array}{ccc} P_1 & P_2 & P_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

The claim is that multiplying these matrices in the order OR, PO, PR produces a personnel-only matrix (OPR) that reflects "the degree of operator interdependence entailed in a given task and personnel structure" where the total number of entries is proportional to the system complexity, the entries along the main diagonal show how autonomous the relevant agent is, and off-diagonals show sources of information in the organization. Perform matrix multiplication in this order to obtain the OPR matrix using transformations as needed where your final matrix has a zero in the last entry of the first row. Which matrix most affects the diagonal values of OPR when it is manipulated?

To preserve this order, the following transformations are necessary: $(OR'PO')'PR$, which is equivalent to $(PO)(OR)(PR)$

$$\left(\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}' \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}' \right)' \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

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

Since this is a linear-additive calculation, multiplying any of the three matrices by a constant has the same effect on the outcome.

3.15 Element-by-element matrix multiplication is a **Hadamard product** (and sometimes called a Schur product), and it is denoted with either "$*$" or "$\odot$" (and occasionally "$\circ$") This element-wise process means that if $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary matrices of identical size, the Hadamard product is $\mathbf{X} \odot \mathbf{Y}$ whose ijth element ($\mathbf{XY}_{ij}$) is $\mathbf{X}_{ij}\mathbf{Y}_{ij}$. It is trivial to see that $\mathbf{X} \odot \mathbf{Y} = \mathbf{Y} \odot \mathbf{X}$ (an interesting exception to general matrix multiplication properties), but show that for two nonzero matrices $\text{tr}(\mathbf{X} \odot \mathbf{Y}) = \text{tr}(\mathbf{X}) \cdot \text{tr}(\mathbf{Y})$. For some nonzero matrix $\mathbf{X}$ what does $\mathbf{I} \odot \mathbf{X}$ do? For an order k $\mathbf{J}$ matrix, is $\text{tr}(\mathbf{J} \odot \mathbf{J})$ different from $\text{tr}(\mathbf{JJ})$? Show why or why not.

$$\text{tr}(\mathbf{X} \odot \mathbf{Y}) = \sum_{i-1}^{k} x_{ii} y_{ii} = \text{sum}(\text{tr}(\mathbf{X})\text{tr}(\mathbf{Y})).$$

turns $\mathbf{X}$ into a diagonal matrix with the same diagonal values.

$J(4)$: $\text{sum}(\text{diag}(J\% * \%J)) = 16,$

$\text{sum}(\text{diag}(J * J)) = 4.$

3.17 Prove that the product of an idempotent matrix is idempotent.

$$\boldsymbol{\beta}'\boldsymbol{\beta} = (\mathbf{A}'\mathbf{A})(\mathbf{A}'\mathbf{A}) = \mathbf{A}'\mathbf{A}\mathbf{A}'\mathbf{A} = \mathbf{A}.$$

3.19 Calculate the LU decomposition of the matrix $\left[\begin{smallmatrix} 2 & 3 \\ 4 & 7 \end{smallmatrix}\right]$ using your preferred software such as with the `lu` function of the `Matrix` library in the `R` environment. Reassemble the matrix by doing the multiplication without using software.

$L = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 4 & 7 \\ 0 & \frac{1}{2} \end{bmatrix}$ From:

```
expand(lu(X))$L %*% expand(lu(X))$U
```

3.21 Demonstrate the inversion property for Kronecker products, $(\mathbf{X} \otimes \mathbf{Y})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1}$, with the following matrices:

$$\mathbf{X} = \begin{bmatrix} 9 & 1 \\ 2 & 8 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 2 & -5 \\ 2 & 1 \end{bmatrix}.$$

$$(\mathbf{X} \otimes \mathbf{Y})^{-1} = \begin{bmatrix} 9\begin{bmatrix} 2 & -5 \\ 2 & 1 \end{bmatrix} & 1\begin{bmatrix} 2 & -5 \\ 2 & 1 \end{bmatrix} \\ 2\begin{bmatrix} 2 & -5 \\ 2 & 1 \end{bmatrix} & 8\begin{bmatrix} 2 & -5 \\ 2 & 1 \end{bmatrix} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 18 & -45 & 2 & -5 \\ 18 & 9 & 2 & 1 \\ 4 & -10 & 16 & -40 \\ 4 & 2 & 16 & 8 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.0095 & 0.0476 & -0.0012 & -0.0060 \\ -0.0190 & 0.0190 & 0.0024 & -0.0024 \\ -0.0024 & -0.0119 & 0.0107 & 0.0536 \\ 0.0048 & -0.0048 & -0.0214 & 0.0214 \end{bmatrix}$$

$$\mathbf{X}^{-1} \otimes \mathbf{Y}^{-1} = \begin{bmatrix} 9 & 1 \\ 2 & 8 \end{bmatrix}^{-1} \otimes \begin{bmatrix} 2 & -5 \\ 2 & 1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.1143 & -0.0143 \\ -0.0286 & 0.1286 \end{bmatrix} \otimes \begin{bmatrix} 0.0833 & 0.4167 \\ -0.1667 & 0.1667 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0095 & 0.0476 & -0.0012 & -0.0060 \\ -0.0190 & 0.0190 & 0.0024 & -0.0024 \\ -0.0024 & -0.0119 & 0.0107 & 0.0536 \\ 0.0048 & -0.0048 & -0.0214 & 0.0214 \end{bmatrix}$$

3.23 For two vectors in $\Re^3$ using $1 = \cos^2\theta + \sin^2\theta$ and $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \mathbf{u}^2 \cdot \mathbf{v}^2$, show that the norm of the cross product between two vectors, $\mathbf{u}$ and $\mathbf{v}$, is: $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\,\|\mathbf{v}\|\sin(\theta)$.

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$$

$$\cos^2(\theta) = \frac{\mathbf{u}^2 \cdot \mathbf{v}^2}{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2}$$

$$1 - \sin^2(\theta) = \frac{\mathbf{u}^2 \cdot \mathbf{v}^2}{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2}$$

$$\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2(\theta) = \mathbf{u}^2 \cdot \mathbf{v}^2$$

$$\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \mathbf{u}^2 \cdot \mathbf{v}^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2(\theta)$$

$$\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2(\theta)$$

$$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \|\mathbf{v}\| \sin(\theta)$$

4

Linear Algebra Continued: Matrix Structure

4.1 For the matrix $\left[\begin{smallmatrix} 3 & 5 \\ 2 & 0 \end{smallmatrix}\right]$, show that the following vectors are or are not in the column space:

$$\begin{bmatrix} 11 \\ 4 \end{bmatrix}, \qquad \begin{bmatrix} 11 \\ 5 \end{bmatrix}.$$

The first vector is in the column space:

$$2\begin{bmatrix} 3 \\ 2 \end{bmatrix} + 1\begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 11 \\ 4 \end{bmatrix},$$

but the second is not because the only way to get 5 in second cell is to use 2.5 as the first multiplier:

$$2.5\begin{bmatrix} 3 \\ 2 \end{bmatrix} + 1\begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 12.5 \\ 5 \end{bmatrix},$$

4.3 Obtain the determinant and trace of the following matrix. Think about tricks to make the calculations easier.

$$\begin{bmatrix} 6 & 6 & 1 & 0 \\ 0 & 4 & 0 & 1 \\ 4 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Since the transpose of a matrix has the same determinant we can use the transpose to put two zeros on the top row. That immediately cuts our efforts by one half since the sub-matrices attached to these values contribute zero to the determinant. Therefore:

$$\begin{vmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 6 & 4 & 2 & 1 \\ 6 & 0 & 4 & 1 \end{vmatrix} = (-1)(1)\begin{vmatrix} 1 & 1 & 0 \\ 6 & 2 & 1 \\ 6 & 4 & 1 \end{vmatrix} + (+1)(1)\begin{vmatrix} 1 & 0 & 0 \\ 6 & 4 & 1 \\ 6 & 0 & 1 \end{vmatrix}$$

where we need to be careful to keep track of the positive and negative multipliers. Fortunately there are three zeros across the top rows of the two determinant calculations so we only have three 2×2 determinant calculations to perform:

$$= (-1)(+1)\begin{vmatrix} 2 & 1 \\ 4 & 1 \end{vmatrix} + (-1)(-1)\begin{vmatrix} 6 & 1 \\ 6 & 1 \end{vmatrix} + (+1)(+1)\begin{vmatrix} 4 & 0 \\ 1 & 1 \end{vmatrix}$$

$$= (-1)(2-4) + (+1)(6-6) + (+1)(4-0) = 6$$

4.5 In their formal study of models of group interaction, Bonacich and Bailey (1971) looked at linear and nonlinear systems of equations (their interest was in models that include factors such as free time, psychological compatibility, friendliness, and common interests). One of their conditions for a stable system was that the determinant of the matrix

$$\begin{pmatrix} -r & a & 0 \\ 0 & -r & a \\ 1 & 0 & -r \end{pmatrix}$$

must have a positive determinant for values of r and a. What is the arithmetic relationship that must exist for this to be true.

$$\det = (+1)(-r)((-r)^2 - 0) + (-1)(a)(0 - a)$$

$$= -r^3 + a^2 > 0 \;\Rightarrow\; a^2 > r^3$$

4.7 Calculate $|B|$, $\text{tr}(B)$, and B^{-1} given $B = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

$$\det(B) = |B| = \begin{vmatrix} 1 & 0 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{vmatrix}, \text{determinant: } 6$$

$$\text{tr}(B) = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \text{trace: } 6$$

$$B^{-1} = \begin{bmatrix} 1 & 0 & -1.33 \\ 0 & 0.5 & 0.00 \\ 0 & 0 & 0.33 \end{bmatrix}^{-1}.$$

4.9 Prove that the following matrix is or is not orthogonal:

$$\mathbf{B} = \begin{bmatrix} 1/3 & 2\sqrt{2}/3 & 0 \\ 2/3 & -\sqrt{2}/6 & \sqrt{2}/2 \\ -2/3 & \sqrt{2}/6 & \sqrt{2}/2 \end{bmatrix}.$$

$\mathbf{B}^{-1} = \mathbf{B}' \;\longrightarrow\; \mathbf{B}'\mathbf{B} = \mathbf{I}$. So:

$$\begin{bmatrix} 1/3 & 2\sqrt{2}/3 & 0 \\ 2/3 & -\sqrt{2}/6 & \sqrt{2}/2 \\ -2/3 & \sqrt{2}/6 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 & -2/3 \\ 2\sqrt{2}/3 & -\sqrt{2}/6 & \sqrt{2}/6 \\ 0 & \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4.11 Clogg, Petkova, and Haritou (1995) give detailed guidance for deciding between different linear regression models using the same data. In this work they define the matrices $\mathbf{X}$, which is $n \times (p+1)$ rank $p+1$, and $\mathbf{Z}$, which is $n \times (q+1)$ rank $q+1$, with $p < q$. They calculate the matrix $A = \left[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}$. Find the dimension and rank of A.

First, the dimension:

$$\Big[\underset{(p+1)\times n}{\mathbf{X}'}\underset{n\times(p+1)}{\mathbf{X}} - \underset{(p+1)\times n}{\mathbf{X}'}\underset{n\times(q+1)}{\mathbf{Z}}(\underset{(q+1)\times n}{\mathbf{Z}'}\underset{n\times q+1)}{\mathbf{Z}})^{-1}$$
$$\times \underset{(q+1)\times n}{\mathbf{Z}'}\underset{n\times(p+1)}{\mathbf{X}}\Big]^{-1}$$

Therefore the dimension of A is $(p+1)\times(p+1)$. The rank of $\mathbf{X}'\mathbf{X}$ is $\min(p+1,n) = p+1$, and the rank of $\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ is $\min(p+1, q+1, n) = p+1$ (i.e. matrix multiplication cannot increase rank). So the rank of the difference is obvious $p+1$ and the matrix is not invertible unless it is full rank, so A is rank $p+1$.

4.13 Land (1980) develops a mathematical theory of social change based on a model of underlying demographic accounts. The corresponding population mathematical models are shown to help identify and track changing social indicators, although no data are used in the article. Label L_x as the number of people in a population that are between x and $x+1$ years old. Then the square matrix $\mathbf{P}'$ of order $(\omega+1)\times(\omega+1)$ is given by

$$\mathbf{P}' = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots \\ L_1/L_0 & 0 & 0 & 0 & \dots \\ 0 & L_2/L_1 & 0 & 0 & \dots \\ 0 & 0 & L_2/L_1 & 0 & \dots \\ \vdots & \vdots & \ddots & \dots & \dots \\ 0 & 0 & 0 & L_\omega/L_{\omega-1} & 0 \end{bmatrix},$$

where ω is the assumed maximum lifespan and each of the nonzero ratios gives the proportion of people living to the next age. The matrix $(\mathbf{I}-\mathbf{P}')$ is theoretically important. Calculate its trace and inverse. The inverse will be a lower triangular form with survivorship probabilities as the nonzero values, and the column sums are standard life expectations in the actuarial sense.

The matrix $(\mathbf{I}-\mathbf{P}')$ is easy to construct:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ -L_1/L_0 & 1 & 0 & 0 & \dots \\ 0 & -L_2/L_1 & 1 & 0 & \dots \\ 0 & 0 & -L_2/L_1 & 1 & \dots \\ \vdots & \vdots & \ddots & \dots & \dots \\ 0 & 0 & 0 & -L_\omega/L_{\omega-1} & 1 \end{bmatrix}.$$

So the trace is obviously one. We can get the inverse from Gauss-Jordan elimination. Start with

$$\left[\begin{array}{ccccc|cccccc} 1 & 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & & \dots \\ -L_1/L_0 & 1 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & & \dots \\ 0 & -L_2/L_1 & 1 & 0 & \dots & 0 & 0 & 1 & 0 & & \dots \\ 0 & 0 & -L_2/L_1 & 1 & \dots & 0 & 0 & 0 & \ddots & & \dots \\ \vdots & \vdots & \ddots & \dots & \dots & \vdots & \vdots & \dots & \dots & & \dots \\ 0 & 0 & 0 & -L_\omega/L_{\omega-1} & 1 & 0 & 0 & 0 & 0 & & 1 \end{array}\right].$$

The first row is already done, so add L_1/L_0 times the first row to the second row:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & -L_2/L_1 & 1 & 0 & \dots \\ 0 & 0 & -L_2/L_1 & 1 & \dots \\ \vdots & \vdots & \ddots & \dots & \dots \\ 0 & 0 & 0 & -L_\omega/L_{\omega-1} & 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ L_1/L_0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & \ddots & \dots \\ \vdots & \vdots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now add L_2/L_1 times row 2 to row 3:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & -L_2/L_1 & 1 & \dots \\ \vdots & \vdots & \ddots & \dots & \dots \\ 0 & 0 & 0 & -L_\omega/L_{\omega-1} & 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ L_1/L_0 & 1 & 0 & 0 & \dots \\ (L_2/L_1)(L_1/L_0) & L_2/L_1 & 1 & 0 & \dots \\ 0 & 0 & 0 & \ddots & \dots \\ \vdots & \vdots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

where $(L_2/L_1)(L_1/L_0)$ simplifies to L_2/L_0. We can continue this way with L_0 cascading down the first row, L_1 down the second row, and so on.

The final line of the matrix inverse will look like:

$$[L_\omega/L_0, L_\omega/L_1, L_\omega/L_2, \dots, L_\omega/L_{\omega-1}, 1].$$

4.15 Consider the two matrices

$$\mathbf{X_1} = \begin{bmatrix} 5 & 2 & 5 \\ 2 & 1 & 2 \\ 3 & 2 & 3 \\ 2.95 & 1 & 3 \end{bmatrix} \qquad \mathbf{X_2} = \begin{bmatrix} 5 & 2 & 5 \\ 2 & 1 & 2 \\ 3 & 2 & 3 \\ 2.99 & 1 & 3 \end{bmatrix}.$$

Given how similar these matrices are to each other, why is $(\mathbf{X_2}'\mathbf{X_2})^{-1}$ so different from $(\mathbf{X_1}'\mathbf{X_1})^{-1}$?

The matrix $(\mathbf{X_2}'\mathbf{X_2})$ is barely nonsingular and therefore presents an unstable matrix inversion:

$$(\mathbf{X2}'\mathbf{X_2})^{-1} = \begin{bmatrix} 16111.11111 & -88.888889 & -16061.11111 \\ -88.88889 & 2.111111 & 87.88889 \\ -16061.11111 & 87.888889 & 16011.61111 \end{bmatrix}.$$

Also, the ratio of smallest to largest eigen value of $\mathbf{X_2}'\mathbf{X_2}$,

$$(103.3808, 0.5593058, 0.00003113)$$

is dramatic: 3320905, compared to 130965.2 for $\mathbf{X_2}'\mathbf{X_2}$ (i.e. 25 times larger).

4.17 A Hilbert matrix has elements $x_{ij} = 1/(i+j-1)$ for the entry in row i and column j. Is this always a symmetric matrix? Is it always positive definite?

As an example, here is a 3×3 Hilbert matrix:

$$H_{3\times 3} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

So it is easy to see from this example and easy to assert from $i+j=j+i$ that the Hilbert matrix of any size is symmetric. For the Hilbert matrix to be positive definite, it must be true that:

$$\mathbf{x}'\mathbf{H}\mathbf{x} > 0, \qquad \forall \mathbf{x} \neq \mathbf{0}$$

for $\mathbf{H}$ an $n \times n$ matrix and $\mathbf{x}$ a n-length vector. Pick an arbitrary column of $\mathbf{H}$ in the operation, j, from $j < n$. The first operation of the product on this column vector is described by:

$$[\mathbf{x}_1\mathbf{H}_{1j} + \mathbf{x}_2\mathbf{H}_{2j} + \ldots + \mathbf{x}_n\mathbf{H}_{nj}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Thus multiplying the row vector here times the column vector gives the jth value in the sum. Note that all values H_{ij} here for $i = 1, \ldots, n$ are positive by definition of the matrix. The corresponding x_i can be either positive or negative in which case $x_i H_{ij} x_i$ is necessarily positive. Therefore the sum down the column j is always positive. Since j was picked arbitrarily from $\mathbf{H}$, the expression above must always be great than zero except for a vector $\mathbf{x}$ of all zeros. Hence the Hilbert matrix of any size is positive definite.

There is an interesting article on just the Hilbert matrix: "Tricks or Treats with the Hilbert Matrix." by Man-Duen Choi (*American Mathematical Monthly* 90 (May, 1983), 301-312).

4.19 Solve the following systems of equations for x, y, and z:

First problem:

Statement of problem:

$$x + y + 2z = 2$$
$$3x - 2y + z = 1$$
$$y - z = 3$$

Multiply 1*third row added to second row and 2*third added to first:

$$x + 3y = 8$$
$$3x - y = 4$$
$$y - z = 3$$

-3*first row added to second row:

$$x + 3y = 8$$
$$-10y = -20 \longrightarrow y = 2$$
$$y - z = 3 \longrightarrow z = -1$$

Final substitution:

$$x + 3(2) = 8 \longrightarrow x = 2$$

Second problem:

Statement of problem:

$$2x + 3y - z = -8$$
$$x + 2y - z = 2$$
$$-x - 4y + z = -6$$

Add third row to first and second rows:

$$x - y = -14$$
$$-2y = -4 \longrightarrow y = 2$$
$$-x - 4y + z = -6$$

Substituting:

$$x - 2 = -14 \longrightarrow x = -12$$
$$-(-12) - 4(2) + z = -6 \longrightarrow z = -10$$

Third problem:

Statement of problem:

$$x - y + 2z = 2$$
$$4x + y - 2z = 10$$
$$x + 3y + z = 0$$

Add row 1 to row 2:

$$x - y + 2z = 2$$
$$5x = 12 \longrightarrow x = \frac{12}{5}$$
$$x + 3y + z = 0$$

Plug in $x = 2$ and add row 1 *3 to row 3:

$$\left(\frac{12}{5}\right) - y + 2z = 2$$
$$4\left(\frac{12}{5}\right) + 7z = 6$$

$$z = -\frac{18}{35} \qquad y = -\frac{22}{35}$$

4.21 A matrix is an M-matrix if $x_{ij} \leq 0, \ \forall i \neq j$, and all the elements of the inverse (X^{-1}) are nonnegative. Construct an example.

The matrix: $\begin{bmatrix} 2.3 & -1.4 \\ -0.3 & 0.4 \end{bmatrix}$ has the inverse: $\begin{bmatrix} 0.8 & 2.8 \\ 0.6 & 4.6 \end{bmatrix}$.

4.23 This question uses the following 8×8 matrix $\mathbf{X}$ of fiscal data by country:

	$\mathbf{x}_{.1}$	$\mathbf{x}_{.2}$	$\mathbf{x}_{.3}$	$\mathbf{x}_{.4}$	$\mathbf{x}_{.5}$	$\mathbf{x}_{.6}$	$\mathbf{x}_{.7}$	$\mathbf{x}_{.8}$
Australia	3.3	9.9	5.41	5.57	5.15	5.35	5.72	6.24
Britain	5.8	11.4	4.81	4.06	4.48	4.59	4.79	5.24
Canada	12.1	9.9	2.43	2.24	2.82	4.29	4.63	5.65
Denmark	12.0	12.5	2.25	2.15	2.42	3.66	4.26	5.01
Japan	4.1	2.0	0.02	0.03	0.10	1.34	1.32	1.43
Sweden	2.2	4.9	1.98	2.55	2.17	3.65	4.56	2.20
Switzerland	−5.3	1.2	0.75	0.24	0.82	2.12	2.56	2.22
USA	5.4	6.2	2.56	1.00	3.26	4.19	4.19	5.44

where $\mathbf{x}_{.1}$ is percent change in the money supply a year ago (narrow), $\mathbf{x}_{.2}$ is percent change in the money supply a year ago (broad), $\mathbf{x}_{.3}$ is the 3-month money market rate (latest), $\mathbf{x}_{.4}$ is the 3-month money market rate (1 year ago), $\mathbf{x}_{.5}$ is the 2-year government bond rate, $\mathbf{x}_{.6}$ is the 10-year government bond rate (latest), $\mathbf{x}_{.7}$ is the 10-year government bond rate (1 year ago), and $\mathbf{x}_{.8}$ is the corporate bond rate (source: *The Economist*, January 29, 2005, page 97). We would expect a number of these figures to be stable over time or to relate across industrialized democracies. Test whether this makes the matrix $\mathbf{X}'\mathbf{X}$ ill-conditioned by obtaining the condition number. What is the rank of $\mathbf{X}'\mathbf{X}$. Calculate the determinant using eigenvalues. Do you expect near collinearity here?

Students should be encouraged to do this problem with software (unless the instructor wants considerable repetition). The following `R` code sets up the problem:

```
fiscal <- matrix(c(3.3,9.9,5.41,5.57,5.15,
          5.34,5.72,6.24,5.8,11.4,4.81,4.06,
          4.48,4.59,4.79,5.24,12.1,9.9,2.43,
          2.24,2.82,4.29,4.63,5.65,12.0,12.5,
          2.25,2.15,2.42,3.66,4.26,5.01,4.1,
```

```
             2.0,0.02,0.03,0.10,1.34,1.32,1.43,
             2.2,4.9,1.98,2.55,2.17,3.65,4.56,
             2.20,-5.3,1.2,0.75,0.24,0.82,2.12,
             2.56,2.22,5.4,6.2,2.56,1.00,3.26,
             4.19,4.19,5.44), byrow=T, ncol=8)
eigen(t(fiscal)%*%fiscal)
svd(t(fiscal)%*%fiscal)
```

This tells us that the condition number is $1.447553e+03/1.296827e-03 = 1116227$, which is quite large suggesting some ill-conditioning. There are no zero eigen values so $\mathbf{X}'\mathbf{X}$ is full rank. The product of the eigenvalues gives the determinant: 7008.848.

4.25 Another method for solving linear systems of equations of the form $\mathbf{A}^{-1}\mathbf{y} = \mathbf{x}$ is **Cramer's rule**. Define $\mathbf{A}_j$ as the matrix where $\mathbf{y}$ is plugged in for the jth column of $\mathbf{A}$. Perform this for every column $1, \ldots, q$ to produce q of these matrices, and the solution will be the vector $\left[\frac{|\mathbf{A}_1|}{|\mathbf{A}|}, \frac{|\mathbf{A}_2|}{|\mathbf{A}|}, \ldots \frac{|\mathbf{A}_q|}{|\mathbf{A}|}\right]$. Show that performing these steps on the matrix in the example on page 159 gives the same answer.

Using the matrix and vector:

$$\mathbf{A} = \begin{bmatrix} 2 & -3 \\ 5 & 5 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

The determinant of $\mathbf{A}$ is $|\mathbf{A}| = 10-(-15) = 25$. Produce the $\mathbf{A}_j$ according to:

$$\mathbf{A}_1 = \begin{bmatrix} 4 & -3 \\ 3 & 5 \end{bmatrix} \qquad \det(\mathbf{A}_1) = 20 - (-9) = 29$$

$$\mathbf{A}_2 = \begin{bmatrix} 2 & 4 \\ 5 & 3 \end{bmatrix} \qquad \det(\mathbf{A}_2) = 6 - 20 = -14.$$

Therefore the inverse with Cramer's rule is:

$$\left[\frac{|\mathbf{A}_1|}{|\mathbf{A}|}, \frac{|\mathbf{A}_2|}{|\mathbf{A}|}\right] = \left[\frac{29}{25}, \frac{-14}{25}\right] = [1.16, -0.56]$$

5

Elementary Scalar Calculus

5.1 Find the following finite limits:

$$\lim_{x\to 4}\left[x^2 - 6x + 4\right] \qquad \lim_{x\to 0}\left[\frac{x-25}{x+5}\right]$$

$$\lim_{x\to 4}\left[\frac{x^2}{3x-2}\right] \qquad \lim_{y\to 1}\left[\frac{y^4-1}{y-1}\right].$$

Finite Limits. Since we know all of these functions are continuous at $x = a$ or $y = a$, we simply solve for $f(a)$.

$$\lim_{x\to 4}(x^2 - 6x + 4) = (4^2 - 6(4) + 4) = -4$$

$$\lim_{x\to 4}\left(\frac{x^2}{3x-2}\right) = \left(\frac{4^2}{3(4)-2}\right) = \frac{8}{5}$$

$$\lim_{x\to 0}\left(\frac{x-25}{x+5}\right) = \left(\frac{0-25}{0+5}\right) = -5$$

$$\lim_{y\to 1}\left(\frac{y^4-1}{y-1}\right) = \lim_{y\to 1}\frac{(y-1)(y+1)(y^2+1)}{(y-1)} = 4$$

5.3 Find the following infinite limits and graph:

$$\lim_{x\to\infty}\left[\frac{9x^2}{x^2+3}\right] \qquad \lim_{x\to\infty}\left[\frac{3x-4}{x+3}\right] \qquad \lim_{x\to\infty}\left[\frac{2^x-3}{2^x+1}\right].$$

Infinite Limit. We can easily solve for each of these infinite limits using L'Hospital's rule, which is switching from $\lim_{x\to\infty}\left(\frac{f(x)}{g(x)}\right)$ to $\lim_{x\to\infty}\left(\frac{f(x)'}{g(x)'}\right)$.

$$\lim_{x\to\infty}\left(\frac{9x^2}{x^2+3}\right) = \lim_{x\to\infty}\left(\frac{18x}{2x}\right) = \lim_{x\to\infty}(9) = 9$$
$$\lim_{x\to\infty}\left(\frac{3x-4}{x+3}\right) = \lim_{x\to\infty}\left(\frac{3}{1}\right) = 3$$
$$\lim_{x\to\infty}\left(\frac{2^x-3}{2^x+1}\right) = \lim_{x\to\infty}\left(\frac{(\log 2)\,2^x}{(\log 2)\,2^x}\right) = 1$$

5.5 Calculate the following derivatives:

$$\frac{d}{dx}3x^{\frac{1}{3}} \qquad \frac{d}{dt}(14t-7)$$
$$\frac{d}{dy}(y^3+3y^2-12) \qquad \frac{d}{dx}\left(\frac{1}{100}x^{25}-\frac{1}{10}x^{0.25}\right)$$
$$\frac{d}{dx}(x^2+1)(x^3-1) \qquad \frac{d}{dy}(y^3-7)\left(1+\frac{1}{y^2}\right)$$
$$\frac{d}{dy}(y-y^{-1})(y-y^{-2}) \qquad \frac{d}{dx}\left(\frac{4x-12x^2}{x^3-4x^2}\right)$$
$$\frac{d}{dy}\exp[y^2-3y+2] \qquad \frac{d}{dx}\log(2\pi x^2).$$

In order to compute these derivatives we use the power rule, $\frac{d}{dx}x^n = nx^{n-1}$.

$$\frac{d(3x^{\frac{1}{3}})}{dx} = x^{-\frac{2}{3}} = \frac{1}{(\sqrt[3]{x})^2}$$

$$\frac{d\left(y^3+3y^2-12\right)}{dy} = 3y^2+6y$$

$$\frac{d\left(x^2+1\right)\left(x^3-1\right)}{dx} = \frac{d(x^5-x^2+x^3-1)}{dx} = 5x^4-2x+3x^2$$

$$\frac{d\left(y-y^{-1}\right)\left(y-y^{-2}\right)}{dy} = \frac{d\left(y^2+y^{-1}-1+y^{-3}\right)}{dy} = 2y+y^{-2}-3y^{-4}$$

$$\frac{d\exp\left(y^2-3y+2\right)}{dy}=\left(e^{y^2-3y+2}\right)(2y-3)$$

$$\frac{d\,(14t-7)}{dt}=14$$

$$\frac{d\left(\frac{1}{100}x^{25}-\frac{1}{10}x^{0.25}\right)}{dx}=\frac{1}{4}x^{24}-\frac{1}{40}x^{-\frac{3}{4}}$$

$$\frac{d\left(y^3-7\right)\left(1+\frac{1}{y^2}\right)}{dy}=\frac{d(y^3+y-7-\frac{7}{y^2})}{dy}=3y^2+\frac{14}{y^3}+1$$

In order to compute the following derivative we will use the quotient rule $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right]=\frac{g(x)\frac{d}{dx}f(x)-f(x)\frac{d}{dx}g(x)}{g(x)^2}$. In words, the derivative of a quotient is the denominator times the derivative of the numerator minus the numerator times the derivative of the denominator, all of this divided by the denominator squared.

$$\frac{d\left(\dfrac{4x-12x^2}{x^3-4x^2}\right)}{dx}$$

$$=\frac{\left[(x^3-4x^2)(4-24x)\right]-\left[\left(4x-12x^2\right)(3x^2-8x)\right]}{(x^3-4x^2)^2}$$

$$=\frac{12x^4-8x^3+16x^2}{x^6-8x^5+16x^4}$$

Finally we will solve the last derivative using the chain rule (page 107), $\frac{d}{dx}f(g(x))=f'(g(x))g'(x)$.

$$\frac{d\log\left(2\pi x^2\right)}{dx}=\frac{1}{2\pi x^2}(4\pi x)=\frac{2}{x}$$

5.7 Calculate the derivative of the following function using logarithmic differentiation:

$$y=\frac{(2x^3-3)^{\frac{5}{2}}}{(x^2-1)^{\frac{2}{3}}(9x^2-1)^{\frac{1}{2}}}.$$

Beginning with $y=\dfrac{(2x^3-3)^{\frac{5}{2}}}{(x^2-1)^{\frac{2}{3}}(9x^2-1)^{\frac{1}{2}}}$, we instead take the derivative of:

$$\log(y)=\frac{5}{2}\log(2x^3-3)-\frac{2}{3}\log(x^2-1)-\frac{1}{2}\log(9x^2-1)$$

$$\frac{1}{y}\frac{d}{dx}y$$

$$= \frac{5}{2}\left(\frac{1}{2x^3-3}\right)(6x^2) - \frac{2}{3}\left(\frac{1}{x^2-1}\right)(2x) - \frac{1}{2}\left(\frac{1}{9x^2-1}\right)(18x)$$

$$= \Bigg(15x^2(x^2-1)(9x^2-1) - \frac{4}{3}x(2x^3-3)(9x^2-1)$$

$$- 9x(2x^3-3)(x^2-1)\Bigg) \Big/ \big((2x^3-3)(x^2-1)(9x^2-1)\big)$$

$$= \frac{15x^2 - 31x + 63x^3 - \frac{388}{3}x^4 + 93x^6}{(2x^3-3)(x^2-1)(9x^2-1)}$$

Now we multiply both sides by y and simplify..

$$\frac{d}{dx}y = \left(\frac{15x^2 - 31x + 63x^3 - \frac{388}{3}x^4 + 93x^6}{(2x^3-3)(x^2-1)(9x^2-1)}\right)$$

$$\times \left(\frac{(2x^3-3)^{\frac{5}{2}}}{(x^2-1)^{\frac{2}{3}}(9x^2-1)^{\frac{1}{2}}}\right)$$

$$= \frac{15x^2 - 31x + 63x^3 - \frac{388}{3}x^4 + 93x^6}{(2x^3-3)^{\frac{2}{3}}(x^2-1)^{\frac{1}{3}}(9x^2-1)^{\frac{1}{2}}}$$

5.9 Using the limit of Riemann integrals, calculate the following integrals. Show steps.

$$R = \int_2^3 \sqrt{x}dx \qquad\qquad R = \int_1^9 \frac{1}{1+y^2}dy$$

$$R = \int_0^{0.1} x^2 dx \qquad\qquad R = \int_1^9 1dx.$$

- With $R = \int_2^3 \sqrt{x}dx$, the problem looks simple but turns out to be quite difficult since there isn't a way (i.e. trick) to step out of the sum into a closed form. There are various approximations out there, but instead let's rely on R where we can give it arbitrarily large numbers of slices. First, let's set the problem up for a left Riemann integral:

$$R = \frac{1}{h}\left[f(2) + f\left(2+\frac{1}{h}\right) + f\left(2+\frac{2}{h}\right) + \ldots + f\left(2+\frac{h-1}{h}\right)\right]$$

$$= \frac{1}{h}\left[(2)^{\frac{1}{2}} + \left(2+\frac{1}{h}\right)^{\frac{1}{2}} + \left(2+\frac{2}{h}\right)^{\frac{1}{2}} + \ldots + \left(2+\frac{h-1}{h}\right)^{\frac{1}{2}}\right]$$

$$= \frac{1}{h}\sum_{i=0}^{h-1}\left(2+\frac{i}{h}\right)^{\frac{1}{2}}$$

where the ordering on the $\frac{1}{h}$ ensures that we stop at 3 as required. Now let's code this algorithm up and extend h:

```
riemann.sqrt <- function(h) {
    sum <- 0
    for (i in 0:(h-1))
        sum <- sum + sqrt(2 + i/h)
    return(sum/h)
}
```

Now we can run this function with increase h:

```
> riemann.sqrt(10)
[1] 1.562538
> riemann.sqrt(10000)
[1] 1.578468
> riemann.sqrt(100000)
[1] 1.578482
> riemann.sqrt(1000000)
[1] 1.578483
```

where we are clearly converging on the correct value (obtainable easily by integration).

- Use the same logic for $R = \int_1^9 \frac{1}{1+y^2} dy$. First, note that the left Riemann integral can be expressed as:

$$\begin{aligned}
R &= \frac{1}{h}\left[f(1) + f(1+\frac{1}{h}) + f\left(2+\frac{2}{h}\right) + \dots \right. \\
&\qquad \left. + f\left(1+\left(8-\frac{2}{n}\right)\right) + f\left(1+\left(8-\frac{1}{n}\right)\right)\right] \\
&= \frac{1}{h}\left[\frac{1}{1+1} + \frac{1}{1+1+\frac{1}{n}} + \frac{1}{1+1+\frac{2}{n}} + \dots \right. \\
&\qquad \left. + \frac{1}{1+1+8-\frac{2}{n}} + \frac{1}{1+1+8-\frac{1}{n}}\right] \\
&= \frac{1}{h}\left[\frac{1}{2} + \frac{1}{2+\frac{1}{n}} + \frac{1}{2+\frac{2}{n}} + \dots + \frac{1}{2+8-\frac{2}{n}} + \frac{1}{2+8-\frac{1}{n}}\right] \\
&= \frac{1}{h}\sum_{i=0}^{8h-1} \frac{1}{2+\frac{i}{h}}
\end{aligned}$$

which leads to following algorithm:

```
riemann.y <- function(h) {
    sum <- 0
    for (i in 0:(8*(h-1)))
        sum <- sum + 1/(2 + i/h)
    return(sum/h)
}
```

Now we can also run this function with increasing h:

```
  riemann.y(100)
[1] 1.604412
> riemann.y(10000)
[1] 1.609388
> riemann.y(100000)
[1] 1.609433
> riemann.y(1000000)
[1] 1.609437
```

- The tricky part to $R = \int_0^{0.1} x^2 dx$ is the limits. As before using a left Riemann sum makes things easier:

$$R = \lim_{h\to\infty} \frac{1}{h} \left[f(0) + f(\left(0 + \frac{1}{h}\right) + \ldots + f(\left(0 + 0.1 - \frac{1}{h}\right) \right]$$

$$R = \lim_{h\to\infty} \frac{1}{h} \left[(0)^2 + (\left(\frac{1}{h}\right)^2 + \ldots + (\left(0.1 - \frac{1}{h}\right)^2 \right]$$

$$R = \lim_{h\to\infty} \frac{1}{h} \sum_{i=0}^{0.1-\frac{1}{h}} \left(\frac{i}{h}\right)^2$$

$$R = \lim_{h\to\infty} \frac{1}{h^3} \sum_{i=0}^{0.1-\frac{1}{h}} (i)^2$$

$$R = \lim_{h\to\infty} \frac{1}{h^3}\frac{1}{6} \left[(0.1h-1)(0.1h-1+1)(2(0.1h-1)+1)\right]$$

$$R = \lim_{h\to\infty} \frac{1}{h^3}\frac{1}{6} \left[0.002h^3 - 0.03h^2 + 0.1h\right]$$

$$R = \lim_{h\to\infty} \frac{1}{6} \left[0.002 - 0.03/h + 0.1/h^2\right]$$

$$= \frac{1}{6}(0.002) = 0.000333\ldots$$

- The area $R = \int_1^9 1dx$ is just a rectangle so h does not matter:

$$R = \sum_{i=1}^{9} 1 = 9 - 1 = 8$$

5.11 Calculate the area of the following function that lies above the x-axis and over the domain $[-10, 10]$:

$$f(x) = 4x^2 + 12x - 18.$$

In order to find the area of the function $f(x) = (4x^2 + 12x - 18)$, within the domain $[10, -10]$, we must first find the roots of the function in this interval to know where the function lies above the x-axis:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-12 \pm \sqrt{144 + 288}}{8}$$
$$= -4.098076, 1.098076$$

(round to -4.1 and 1.1 for ease if desired). We know that this is a quadratic form open upwards, so we need to integrate outwards from these two points ignoring the interval between:

$$\int_{-10}^{-4.1} (4x^2 + 12x - 18)dx = \frac{4}{3}x^3 + 6x^2 - 18x\bigg|_{-10}^{-4.1}$$
$$= \left(\frac{4}{3}(-4.1)^3 + 6(-4.1)^2 - 18(-4.1)\right)$$
$$- \left(\frac{4}{3}(-10)^3 + 6(-10)^2 - 18(-10)\right)$$
$$= (-118.9547) - (-1753.3330) = 1634.378$$

$$\int_{1.1}^{10} (4x^2 + 12x - 18)dx = \frac{4}{3}x^3 + 6x^2 - 18x\bigg|_{10}^{1.1}$$
$$= \left(\frac{4}{3}(10)^3 + 6(10)^2 - 18(10)\right)$$
$$- \left(\frac{4}{3}(1.1)^3 + 6(1.1)^2 - 18(1.1)\right)$$
$$= (1753.330) - (-10.76533) = 1764.098$$

So the total area under curve above the x-axis between -10 and 10 is $1634.378 + 1764.098 = 3398.476$.

5.13 Calculate the following indefinite integrals:

$$\int (2 + 12x^2)^{3/2} dx \qquad \int (x^2 - x^{-\frac{1}{2}})dx \qquad \int \frac{y}{\sqrt{9 + 3y^3}} dy$$

$$\int 360t^6 dt \qquad \int \frac{x^2}{1 - x^2} dx \qquad \int (11 - 21y^9)^2 dy.$$

$$\int (2+12x^2)^{\frac{3}{2}}dx = \frac{3}{4}x\sqrt{12x^2+2}+$$

$$\frac{1}{4}\sqrt{3}\log\left(2x\sqrt{3}+\sqrt{12x^2+2}\right)$$

$$+x\left(\frac{1}{2}\sqrt{12x^2+2}+3x^2\sqrt{12x^2+2}\right)+c$$

$$\int (x^2 - x^{-\frac{1}{2}})dx = \frac{1}{3}x^3 - 2\sqrt{x}+c$$

$$\int (360t^6)dt = \frac{360}{7}t^7 + c$$

$$\int \left(\frac{x^2}{1-x^2}\right)dx = \frac{1}{2}\log\left(x+1\right) - \frac{1}{2}\log\left(x-1\right) - x + c$$

$$\int \frac{y}{\sqrt{9+3y^3}}dy = \frac{1}{\sqrt{3}}\int \frac{y}{\sqrt{3+y^3}}dy$$

$\int \frac{y}{\sqrt{3+y^3}}dy$ needs to be solved with the reduction formula:

$$\int \frac{x^n}{z_3^m}dx = \frac{x^{n-2}}{z_3^{m-1}(n+1-3m)b} - \frac{(n-2)a}{b(n+1-3m)}\int \frac{x^{n-3}}{z_3^m}dx$$

where: $z_3 = a + bx^3$

so:$n = 1,\ a = 3,\ b = 1,\ m = \frac{1}{2}\ \rightarrow\ z_3 = a + by^3$

the first iteration looks like:

$$\int \frac{y}{\sqrt{3+y^3}}dy = 2y^{-1}(3+y^3)^{\frac{1}{2}} + 6\int \frac{y^{-2}}{(3+y^3)^{\frac{1}{2}}}dy$$

This still contains an integral which is unsatisfactory, but we can evaluate the limit:

$$\int \frac{y}{\sqrt{3+y^3}}dy = \sum_{i=1}^{\infty}\left[\frac{y^{2-3i}}{(3+y^3)^{-\frac{1}{2}}(5-3i-\frac{3}{2}} \right.$$
$$\left. - \frac{12-18i}{10-6i-3}\int \frac{y^{1-3i}}{(3+y^3)}dy\right]$$

$$\int (11-21y^9)^2 dy = 121y - \frac{231}{5}y^{10} + \frac{441}{19}y^{19} + c$$

5.15 For the gamma function (Section 5.6.4.1):

a) Prove that $\Gamma(1) = 1$.

b) Express the binomial function, $\binom{n}{y} = \frac{n!}{y!(n-y)!}$, in terms of the gamma function instead of factorial notation for integers.

c) What is $\Gamma\left(\frac{1}{2}\right)$?

Recall: $\Gamma(\omega) = \int_0^{\infty} t^{\omega-1}e^{-t}dt, \quad \omega > 0.$

$$\int_0^{\infty} t^0 e^{-t}dt = -e^{-t}\Big|_{t=0}^{t=\infty} = 0-(-1) = 1.$$

$$\binom{n}{y} = \frac{n!}{y!(n-y)!} = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}.$$

$$\Gamma(\frac{1}{2}) = \sqrt(\pi).$$

5.17 Blackwell and Girshick (1954) derived the result below in the context of mixed strategies in game theory. Game theory is a tool in the social sciences where the motivations, strategies, and rewards to hypothesized competing actors are analyzed mathematically to make predictions or explain observed behavior. This idea originated formally with von Neumann and Morgenstern's 1944 book. Simplified, an actor employs a mixed strategy when she has a set of alternative actions each with a known or assumed probability of success, and the choice of actions is made by randomly selecting one with that associated probability. Thus if there are three alternatives with success probabilities $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$, then there is a 50% chance of picking the first, and so on. Blackwell and Girshick (p.54) extended this idea to continuously measured alternatives between zero and one (i.e., a smooth function rather than a discrete listing). The first player accordingly chooses value $x \in [0:1]$, and the second player chooses value $y \in [0:1]$, and the function that defines the "game" is given by $M(x, y) = f(x - y)$, where

$$f(t) = \begin{cases} t(1-t), & \text{for } 0 \geq t \geq 1 \\ f(t+1), & \text{for } -1 \geq t \geq 0. \end{cases}$$

In other words, it matters which of x and y is larger in this game. Here is the surprising part. For any fixed value of y (call this y_0), the expected value of the game to the first player is $\frac{1}{6}$. To show this we integrate over a range of alternatives available to this player:

$$\begin{aligned} \int_0^1 M(x, y)dx &= \int_0^1 f(x - y_0)dx \\ &= \underbrace{\int_0^{y_0} f(x - y_0)dx}_{x<y_0} + \underbrace{\int_{y_0}^1 f(x - y_0)dx}_{x>y_0}, \end{aligned}$$

where breaking the integral into two pieces is necessary because the first one contains the case where $-1 \geq t \geq 0$ and the second one contains the case where $0 \geq t \geq 1$. Substitute in the two function values (t or $t+1$) and integrate over x to obtain exactly $\frac{1}{6}$.

$$\begin{aligned}
\int_0^1 M(x,y)dx &= \int_0^{y_0} (x - y_0 + 1)(1 - (x - y_0 + 1))dx \\
&\quad + \int_{y_0}^1 (x - y_0)(1 - (x - y_0))dx \\
&= \int_0^{y_0} (x - y_0 + 1)(1 - (y_0 - x))dx \\
&\quad + \int_{y_0}^1 (x - y_0)(1 - x + y_0)dx \\
&= \int_0^{y_0} (-x^2 + x(2y_0 - 1) + y_0 - y_0^2)dx \\
&\quad + \int_{y_0}^1 (-x^2 + x(2y_0 + 1) - y_0 - y_0^2)dx \\
&= -\frac{1}{3}x^3 + \frac{1}{2}x^2(2y_0 - 1) + y_0 x - y_0^2 x \Big|_0^{y_0} \\
&\quad + -\frac{1}{3}x^3 + \frac{1}{2}x^2(2y_0 + 1) - y_0 x - y_0^2 x \Big|_{y_0}^1 \\
&= -\frac{1}{3}y_0^3 + \frac{1}{2}y_0^2(2y_0 - 1) + y_0^2 - y_0^3 - \frac{1}{3} \\
&\quad + \frac{1}{2}(2y_0 - 1) - y_0 - y_0^2 \\
&\quad - (-\frac{1}{3}y_0^3 + \frac{1}{2}y_0^2(2y_0 + 1) - y_0^2 - y_0^3) \\
&= \frac{1}{6}
\end{aligned}$$

5.19 Show that the Mean Value Theorem is a special case of Rolle's Theorem by generalizing the starting and stopping points.

One can show that either is a special case of the other. To show that Rolle's Theorem is a special case of the MVT, we just have to stipulate that $f(a) = f(b) = 0$. To show the reverse (which is much more interesting), define a new function:

$$g(x) = f(x) - x\frac{f(b) - f(a)}{b - a} + a\frac{(f(b) - f(a))}{b - a} - f(a)$$

This assumes, without losing generality, that $b > a$ and $f(b) > f(a)$. We can relax this assumption by using absolute values, if desired. Now we can show that:

$$
\begin{aligned}
g(a) &= f(a) - a\frac{f(b)-f(a)}{b-a} + a\frac{(f(b)-f(a))}{b-a} - f(a) = 0 \\
g(b) &= f(b) - b\frac{f(b)-f(a)}{b-a} + a\frac{(f(b)-f(a))}{b-a} - f(a) \\
&= f(b) - f(a) - \frac{f(b)-f(a)}{b-a}(b-a) = 0
\end{aligned}
$$

Therefore the endpoints are now both on x-axis. Finally, taking the derivative produces:

$$
g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}
$$

and we know that $f'(x) = \frac{f(b)-f(a)}{b-a}$ from the definition of the MVT, so $g'(x) = 0$ and we're done.

5.21 From the appendix to Krehbiel (2000), take the partial derivative of

$$
\frac{50((M^2 - M/2 + \delta)^2 + 100(1 - M/2 + \delta) - M)}{M((1 - M/2 + \delta) - (M/2 + \delta))}
$$

with respect to M and show that it is decreasing in $M \in (0\!:\!1)$ (i.e., as M increases) for all values $\delta \in (0\!:\!1)$.

$$
\begin{aligned}
f(M) &= \frac{50((M^2 - M/2 + \delta)^2 + 100(1 - M/2 + \delta) - M)}{M((1 - M/2 + \delta) - (M/2 + \delta))} \\
&= 50(M^4 - \frac{1}{2}M^3 + M^2\delta - \frac{1}{2}M^3 - \frac{1}{4}M^2 - \frac{1}{2}M\delta + M^2\delta \\
&\quad - \frac{1}{2}M\delta + \delta^2 + 100 - 50M + 100\delta - M)(M(1-M))^{-1} \\
&= 50(M^4 - M^3 + (2\delta - \frac{1}{4})M^2 \\
&\qquad - (49+\delta)M + 100 + 100\delta + \delta^2)(M^2 - M)^{-1}
\end{aligned}
$$

$$\frac{d}{dM}f(M)$$

$$= 50\Bigg[(4M^3 - 3M2 + 2\left(2\delta - \frac{1}{4}\right)M - (49+\delta))(M^2 - M)^{-1}$$
$$+ (M^4 - M^3 + \left(2\delta - \frac{1}{4}\right)M^2 - (49+\delta)M + 100 + 100\delta + \delta^2)$$
$$\times (-1)(M^2 - M)^{-2}(2M-1)\Bigg]$$

$$= 50\Bigg[\frac{4M^3 - 3M2 + 2(2\delta - \frac{1}{4})M - (49+\delta)}{M^2 - M}$$
$$- \frac{(M^4 - M^3 + (2\delta - \frac{1}{4})M^2 - (49+\delta)M + 100 + 100\delta + \delta^2)(2M-1)}{(M^2-M)^2}\Bigg]$$

$$= \frac{50}{(M^2-M)^2}\Bigg[(4M^3 - 3M2 + 2(2\delta - \frac{1}{4})M - (49+\delta))(M^2 - M)$$
$$- (M^4 - M^3 + (2\delta - \frac{1}{4})M^2 - (49+\delta)M + 100 + 100\delta + \delta^2)(2M-1)\Bigg]$$

$$= \frac{50}{(M^2-M)^2}\Bigg[4M^5 - 3M^4 + \left(4\delta - \frac{1}{2}\right)M^3 - (49+\delta)M^2$$
$$- 4M^4 + 3M^3 - \left(4\delta - \frac{1}{2}\right)M^2 + (49+\delta)M$$
$$- (2M^5 - 2M^4 + \left(4\delta - \frac{1}{2}\right)M^3 - 2(49+\delta)M^2 + 200M + 200M\delta$$
$$+ 2M\delta^2 - M^4 + M^3 - \left(2\delta - \frac{1}{4}\right)a + (49+\delta)M - 100 - 100\delta - \delta^2)\Bigg]$$

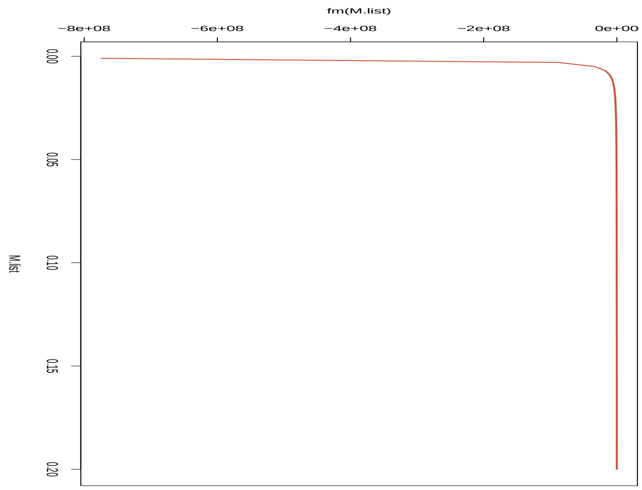
$$= \frac{50}{(M^2-M)^2}\Bigg[2M^5 - 4M^4 + 2M^3 + (-49 - \delta - 4\delta + \frac{1}{2} + 98 + 2\delta + 2\delta$$
$$- \frac{1}{4})M^2(49 + \delta - 200 - 200\delta - 2\delta^2 + 49 - \delta)M - 100 - 100\delta - \delta^2\Bigg]$$

$$= \frac{50}{(M^2-M)^2}\Bigg[2M^5 - 4M^4 + 2M^3 + \left(\frac{99}{2} - \delta\right)M^2$$
$$- (102 + 200\delta + \delta^2)M - 100 - 100\delta - \delta^2\Bigg]$$

$$\begin{aligned}
&\frac{d}{dM}f(M) \\
&= \frac{50(2M^3)(M^2-2M+1)}{M^2(M^2-2M+1)} + \frac{50}{(M^2-M)^2}\left[\left(\frac{99}{2}-\delta\right)M^2\right. \\
&\qquad \left. -(102+200\delta+\delta^2)M - 100 - 100\delta - \delta^2\right] \\
&= 100M + \frac{50}{(M^2-M)^2}\left[\left(\frac{99}{2}-\delta\right)M^2\right. \\
&\qquad \left. -(102+200\delta+\delta^2)M - 100 - 100\delta - \delta^2\right] \\
&= 100M + \frac{50}{(M^2-M)^2}\left[\left(\frac{99}{2}-\delta\right)M^2\right. \\
&\qquad \left. -\left(\frac{99}{2}-\delta\right)M - \left(\frac{105}{2}+201\delta+\delta^2\right)M + 100 + 100\delta + \delta^2\right] \\
&= 100M + \frac{1}{M^2-M}\left(50\left(\frac{99}{2}-\delta\right)\right) \\
&\qquad - \frac{50}{(M^2-M)^2}\left[\left(\frac{105}{2}+201\delta+\delta^2\right)M + 100 + 100\delta + \delta^2\right] \\
&\qquad = 100M + \frac{2475-50\delta}{M^2-M} \\
&\qquad\quad - \frac{(2625+10050\delta+\delta^2)M + 500 + 500\delta + 50\delta^2}{(M^2-M)^2}
\end{aligned}$$

This function rapidly increases in M for any value of θ and then remains at zero. As illustrated:

```
fm <- function(M,delta=0.5)
    100*M + (2475 - 50*delta)/(M^2-M) -
    ((2625 + 10050*delta + delta^2)*M +
    500 + 500*delta + 50*delta^2)/((M^2-M)^2)
M.list <- seq(0.001,0.2,length=100)
plot(M.list,fm(M.list),type="l",
    col="tomato3",lwd=2)
```


fm(M.list)
−8e+08
−6e+08
−4e+08
−2e+08
0e+00
M.list
0.00
0.05
0.10
0.15
0.20

6

Additional Topics in Scalar and Vector Calculus

6.1 For the following functions, determine the interval over which the function is convex or concave (if at all), and identify any inflection points:

$$f(x) = \frac{1}{x} \qquad f(x) = x^3$$

$$f(x) = x^2 + 4x + 8 \qquad f(x) = -x^2 - 9x + 16$$

$$f(x) = \frac{1}{1+x^2} \qquad f(x) = \frac{\exp(x)}{(1+\exp(x)}$$

$$f(x) = 1 - \exp(-\exp(x)) \qquad f(x) = \frac{x^{7/2}}{2+x^2}$$

$$f(x) = (x-1)^4(x+1)^3 \qquad f(x) = \log(x).$$

- $f(x) = \frac{1}{x}$.

$$\frac{d}{dx}f(x) = -x^{-2} \qquad \frac{d^2}{d^2x} = 2x^{-3}$$

  So this function is concave to the x-axis from the negative side. There are no inflection points since the second derivative doesn't change signs.

- $f(x) = x^3$.

$$\frac{d}{dx}f(x) = 3x^2 \qquad \frac{d^2}{dx} = 6x$$

- $f(x) = x^2+4x+8$. This is a quadratic function with a positive coefficient on the square term. Therefore it is convex to the x-axis and has no inflection points.
- $f(x) = -x^2-9x+16$. This is another quadratic, but it faces downward. It crosses the x-axis (has roots) at -10.52079, and 1.52079, so it is concave between these values.
- $f(x) = \frac{1}{1+x^2}$. This function is symmetric around zero and concave between the two inflection points and convex outside of them. These are calculated as follows.

$$\begin{aligned} f^{'}(x) &= -2x(1+x^2)^{-2} \\ f^{''}(x) &= -2(1+x^2)^{-2} + 8x^2(1+x^2)^{-3} \\ &= 2(1+x^2)^{-2}(-1+4x^2(1+x^2)^{-1}) \end{aligned}$$

It is clear from the second form that the second derivative will be equal to zero if $-1+4x^2(1+x^2)^{-1}$ is zero and that $2(1+x^2)^{-2}$ cannot provide a zero term. Therefore the root is found by:

$$\begin{aligned} -1+4x^2(1+x^2)^{-1} &= 0 \\ 4x^2(1+x^2)^{-1} &= 1 \\ 4x^2 &= 1+x^2 \\ x^2 &= 3 \\ x &= \pm\sqrt{3} \end{aligned}$$

which can be confirmed by plugging this value back into the second derivative function:

$$-2\left(1+\frac{1}{3}\right)^{-2} * 8\left(\frac{1}{3}\right)\left(1+\frac{1}{3}\right)^{-3} = 0$$

- The function $f(x) = \frac{\exp(x)}{(1+\exp(x))}$ is of course the logit function, which we know a lot about. The first and second derivatives are given by:

$$f^{'}(x) = \exp(x)[1+\exp(x)]^{-1} + \exp(x)^2[1+\exp(x)]^{-2}$$

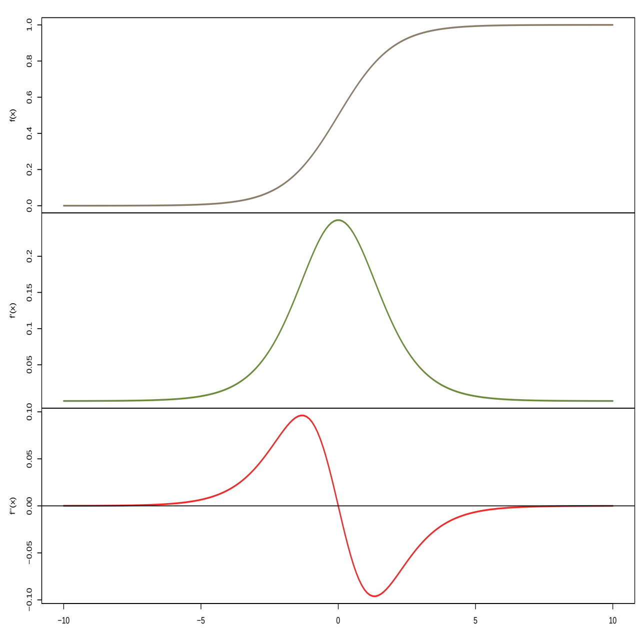
$$\begin{aligned} f^{''}(x) = \exp(x)[1+\exp(x)]^{-1} - 3\exp(x)^2[1+\exp(x)]^{-2} \\ + 2\exp(x)^3[1+\exp(x)]^{-3} \end{aligned}$$

If we plug $x = 0$ into the second derivative function we get:

$$(1)(1+1)^{-1} - 3(1)(1+1)^{-2} + 2(1)(1+1)^{-3} = \frac{1}{2} - \frac{3}{4} + \frac{2}{8} = 0$$

so zero is an inflection point. This can also be seen easily in the graph below. Since zero is an inflection point and $y = 0.5$ is the corresponding y value, then the function is concave (negative second derivative) at $x > 0$ and convex (positive second derivative) at $x < 0$.

Logit Graph (sub-problem 6):



- For $f(x) = 1 - \exp(-\exp(x))$, the first and second derivatives are given by:

$$\begin{aligned} f^{'}(x) &= -\exp(-\exp(x))(-\exp(x)) \\ &= \exp(x)\exp(-\exp(x)) \\ f^{''}(x) &= \exp(x)\exp(-\exp(x)) + \exp(x)\exp(-\exp(x))(-\exp(x)) \\ &= \exp(x)\exp(-\exp(x))(1-\exp(x)) \end{aligned}$$

In the second derivative function $\exp(x)\exp(-\exp(x))$ cannot possibly provide zero values since the log of zero is undefined. So the only way it can be zero is if $(1 - \exp(x)) = 0$, which occurs only at $x = 0$. So this must be an inflection point. Plugging zero back into the original function produces $y = 0.6321206$ at the inflection point. This function is the "cloglog" function used in some regression models as an alternate link function to the logit (above). It is therefore interesting to note that the inflection is *not* 0.5.

- The function $f(x) = \frac{x^{7/2}}{2+x^2}$ has first and second derivatives:

$$\begin{aligned} f^{'}(x) &= \frac{7}{2}x^{\frac{5}{2}}(2+x^2)^{-1} - 2x^{\frac{9}{2}}(2+x^2)^{-2} \\ f^{''}(x) &= \frac{7}{2}\frac{5}{2}x^{\frac{3}{2}}(2+x^2)^{-1} + \frac{7}{2}x^{\frac{5}{2}}(2+x^2)^{-2}(-1)(2x) \\ &\quad -(2)\frac{9}{2}x^{\frac{7}{2}}(2+x^2)^{-2} - 2x^{\frac{9}{2}}(2+x^2)^{-3}(-2)(2x) \\ &= \frac{35}{4}x^{\frac{3}{2}}(2+x^2)^{-1} - 16x^{\frac{7}{2}}(2+x^2)^{-2} - 8x^{\frac{11}{2}}(2+x^2)^{-3} \end{aligned}$$

So there is clearly an inflection point at $x = 0$, and the term $(2+x^2)$ cannot produce one due to the square and that it is always in the denominator. To determine if there are any other inflection points we will set the second derivative equal to zero and attempt to solve for x:

$$0 \equiv \frac{35}{4}x^{\frac{3}{2}}(2+x^2)^{-1} - 16x^{\frac{7}{2}}(2+x^2)^{-2} - 8x^{\frac{11}{2}}(2+x^2)^{-3}$$

$$= 35x^2x^{\frac{3}{2}}(2+x^2)^2 - 64x^{\frac{7}{2}}(2+x^2) + 32x^{\frac{11}{2}}$$

$$= 34(4+4x^2+x^4) - 128x^2 - 64x^4 + 32x^4$$

$$= 140 + 12x^2 + 3x^4$$

Substitute $y = x^2$ since a root in y implies a root in x:

$$0 = 3y^2 + 12y + 140 = (3y+3)(y+3) + 131$$

but when we apply the quadratic formula the square root component gives $\sqrt{-1536}$, meaning there is no root in y so there is no root in x. Therefore the only inflection point is the previously found one at zero. Since small positive values of x give a positive second derivative function and positive values of the original function, then the function is concave above zero. By the same logic, since small negatives values of x give a negative second derivative function value and the original function is also negative, then the function is concave to the x-axis from the negative side for values of x less than zero.

- For the function $f(x) = (x-1)^4(x+1)^3$, we have:

$$f^{'}(x) = 4(x-1)^3(x+1)^3 + 3(x-1)^4(x+1)^2$$

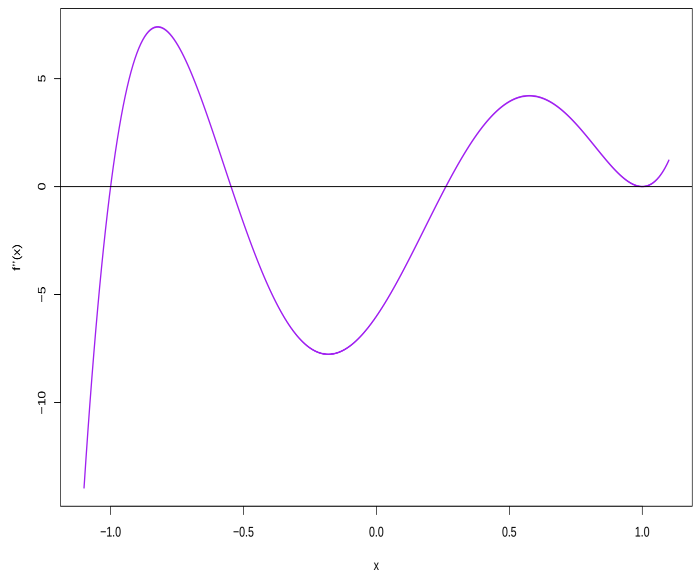
$$f^{''}(x) = 12(x-1)^2(x+1)^3 + 4(x-1)^3(x+1)^2(3)$$
$$+ 4(x-1)^3(x+1)^2)(3) + (x-1)^4(x+1)(2)(3)$$

$$= 12(x-1)^2(x+1)^3 + 24(x-1)^3(x+1)^2$$
$$+ 6(x-1)^4(x+1)$$

Clearly the second derivative has two and only two roots at $+1$ and -1. The second derivative, with some mechanical algebra work, can be expressed as:

$$f^{''}(x) = 42x^5 - 30x^4 - 60x^3 + 36x^2 + 18x - 6$$

and so we have no way of knowing that we have captured all the roots. Plotting the function (see below) indicates two more near zero.



Analytically these can be hard to find, but we can use Newton-Raphson to save work. Suppose we notate $f^{''}(x) = g(x)$, then

$$g^{'}(x) = 210x^4 - 120x^3 - 180x^2 + 72x + 18$$

For starting points we can look at the graph. Here is some `R` code to find the roots:

```
x.vals <- -0.5 # or x.vals <- -0.5
for (i in 1:9)  {
    x <- x.vals[i]
    x.vals <- c( x.vals,x -
```

```
            (42*x^5-30*x^4-60*x^3+36*x^2+18*x-6)/
            (210*x^4-120*x^3-180*x^2+72*x+18) )
}
```

This produces the following tabulated results:

Iteration	Run 1	Run 2
1	0.3000000	-0.5000000
2	0.2597887	-0.5483871
3	0.2612025	-0.5469188
4	0.2612039	-0.5469182
5	0.2612039	-0.5469182
6	0.2612039	-0.5469182
7	0.2612039	-0.5469182
8	0.2612039	-0.5469182
9	0.2612039	-0.5469182
10	0.2612039	-0.5469182

We know that there are not any additional roots because the term $52x^5$ quickly dominates outside of $[-1:1]$. Since it has an odd power, then the function head to infinity in either direction as implied by the graph. Also from the graph we see concavity in $[-1:-0.54]$ and $[0.26:1]$ with concavity from the negative side in $[-0.54:0.26]$.

- $f(x) = \log(x)$. We know that the log function is concave above the x-axis starting at the only root $x = 1$.

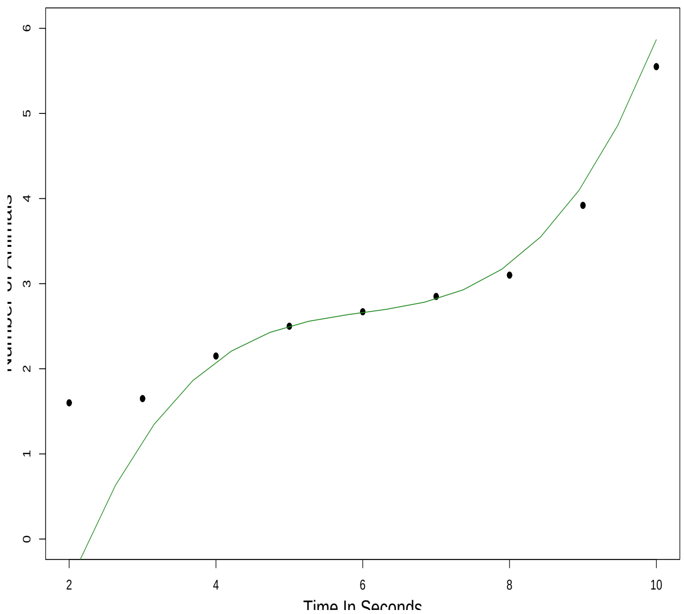
6.3 To test memory retrieval, Kail and Nippold (1984) asked 8-, 12-, and 21-year-olds to name as many animals and pieces of furniture as possible in separate 7–minute intervals. They found that this number increased across the tested age range but that the rate of retrieval slowed down as the period continued. In fact, the responses often came in “clusters” of related responses (“lion,” “tiger,” “cheetah,” etc.), where the relation of time in seconds to cluster size was fitted to be $cs(t) = at^3 + bt^2 + ct + d$, where time is t, and the others are estimated parameters (which differ by topic, age group, and subject). The researchers were very interested in the inflection point of this function because it suggests a change of cognitive process. Find it for the unknown parameter values analytically by taking the second derivative. Verify that it is an inflection point and not a maxima or minima. Now graph this function for the points supplied in the authors' graph of one particular case for an 8-year-old: $cs(t) = [1.6, 1.65, 2.15, 2.5, 2.67, 2.85, 3.1, 4.92, 5.55]$ at the points $t = [2, 3, 4, 5, 6, 7, 8, 9, 10]$. They do not give parameter values for this case, but plot the function on the same graph for the values $a = 0.04291667$, $b = -0.7725$, $c = 4.75$, and $d = -7.3$. Do these values appear to satisfy your result for the inflection point?

$$
\begin{aligned}
cs(t) &= at^3 + bt^2 + ct + d \\
\frac{d}{dt}cs(t) &= 3at^2 + 2bt + t \\
\frac{d^2}{d^2t}cs(t) &= 6at + 2b
\end{aligned}
$$

Setting this equal to zero and solving gives: $t^{'} = -\frac{b}{3a}$. Since the function is smooth, this is the point where the second derivative function changes sign, meaning an inflection point. Since a is in the denominator and cannot be zero, there is no inflection point for this function when there is no third order term.

The points appear to fit the functional form except at the ends.

6.5 Smirnov and Ershov (1992) chronicled dramatic changes in public opinion during the period of "Perestroika" in the Soviet Union (1985 to 1991). They employed a creative approach by basing their model on the principles of thermodynamics with the idea that sometimes an encapsulated liquid is immobile and dormant and sometimes it becomes turbulent and pressured,



literally letting off steam. The catalyst for change is hypothesized to be radical economic reform confronted by conservative counter-reformist policies. Define p as some policy on a metric $[-1\!:\!1]$ representing different positions over this range from conservative ($p < 1$) to liberal ($p > 1$). The resulting public opinion support, S, is a function that can have single or multiple modes over this range, inflection points and monotonic areas, where the number and variety of these reflect divergent opinions in the population. Smirnov and Ershov found that the most convenient mathematical form here was

$$S(p) = \sum_{i=1}^{4} \lambda_i p^i,$$

where the notation on p indicates exponents and the λ_i values are a series of specified scalars. Their claim was that when there are two approximately equal modes (in $S(p)$), this represents the situation where "the government ceases to represent the majority of the electorate." Specify λ_i values to give this shape; graph over the domain of p; and use the first and second derivatives of $S(p)$ to identify maxima, minima, and inflection points.

Given that the support covers both sides of zero, we know that a dominant quadratic form will give us two modes. We can get a tapering off from the modes by manipulating the fourth exponent to be negative. We also know that the first term is not particularly important since it just scales upward. An interesting form within the parameters of the problem is given by $\boldsymbol{\lambda} = [0, 20, 1, -12]$. The first derivative of $S(p)$ is:

$$\frac{d}{dp} = 40p + 2p^2 - 48p^3$$

$$= p(-48p^2 + 2p + 40)$$

So one root is obviously at zero. The other two are found using the quadratic formula:

$$p = \frac{-2 \pm \sqrt{4 - 4(-48)(40)}}{2(-48)}$$

$$= 0.933942 \text{ or } -0.8922753$$

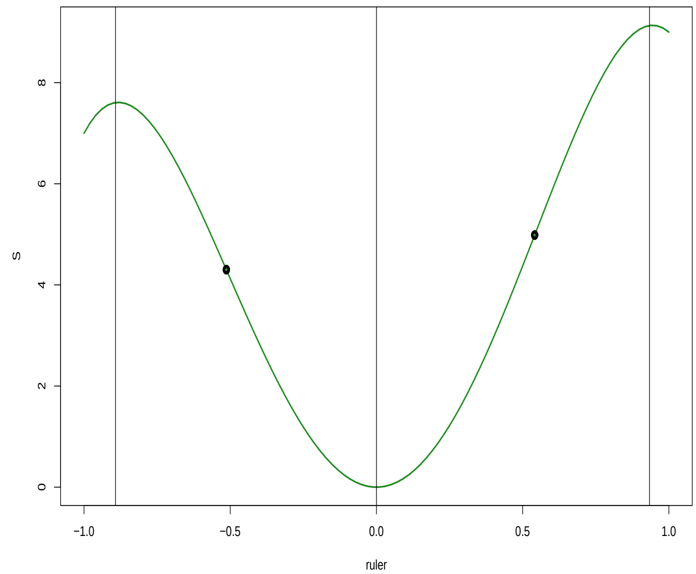
This form is graphed here with the roots indicated by vertical lines.

The second derivative of function is:

$$\frac{d}{dp}\left(-48p^3 + 2p^2 = -144p^2 + 4p + 40\right)$$

Setting this equal to zero and solving the quadratic produces the two points $(0.5411181, -0.5133404)$. Since these are the points where the second derivative is zero and the function is not a maximum or minimum, then we know that the derivative changes sign on either side of these points and therefore they are inflection points. The graph shows these points on the original function.

While the chosen function is arbitrary, students can easily pick their own version and obtain the points of interest. To be consistent with the article, the function should have two maxima, which also makes the analysis more interesting since it produces three roots and two inflection points.



6.7 For the function $f(u, v) = \sqrt{u + v^2}$, calculate the partial derivatives with respect to u and v and provide the value of these functions at the point $(\frac{1}{2}, \frac{1}{3})$.

- $\frac{\partial}{\partial u} f(u, v) = \frac{1}{2}(u + v^2)^{-\frac{1}{2}}$
- $\frac{\partial}{\partial u} f(\frac{1}{2}, \frac{1}{3}) = 0.6396021$
- $\frac{\partial}{\partial v} f(u, v) = v(u + v^2)^{-\frac{1}{2}}$
- $\frac{\partial}{\partial v} f(\frac{1}{2}, \frac{1}{3}) = 0.4264014$

6.9 Obtain the first, second, and third derivatives of the following functions:

$$f(x) = 5x^4 + 3x^3 - 11x^2 + x - 7 \qquad h(z) = 111z^3 - 121z$$

$$f(y) = \sqrt{y} + \frac{1}{y^{\frac{7}{2}}} \qquad f(x) = (x^9)^{-2}$$

$$h(u) = \log(u) + k \qquad g(z) = \sin(z) - \cos(z).$$

$$f(x) = 5x^4 + 3x^3 - 11x^2 + x - 7$$
$$f'(x) = 20x^3 + 9x^2 - 22x + 1$$
$$f''(x) = 60x^2 + 18x - 22$$
$$f'''(x) = 120x + 18$$

$$h(z) = 111z^3 - 121z$$
$$h'(z) = 333z^2 - 121$$
$$h''(z) = 666z$$
$$h'''(z) = 666$$

$$f(y) = y^{\frac{1}{2}} + y^{-\frac{7}{2}}$$
$$f'(y) = \frac{1}{2}y^{-\frac{1}{2}} - \frac{7}{2}y^{-\frac{9}{2}}$$
$$f''(y) = -\frac{1}{4}y^{-\frac{3}{2}} + \frac{63}{4}y^{-\frac{11}{2}}$$
$$f'''(y) = \frac{3}{8}y^{-\frac{5}{2}} - \frac{693}{8}y^{-\frac{13}{2}}$$

$$f(x) = x^{-18}$$
$$f'(x) = -18x^{-19}$$
$$f''(x) = 342x^{-20}$$
$$f'''(x) = -6840x^{-21}$$

$$\begin{aligned} h(u) &= \log u + k \\ h'(u) &= u^{-1} \\ h''(u) &= -u^{-2} \\ h'''(u) &= 2u^{-3} \end{aligned}$$

$$\begin{aligned} g(z) &= \sin(z) - \cos(z) \\ g'(z) &= \cos(z) + \sin(z) \\ g''(z) &= -\sin(z) + \cos(z) \\ g'''(z) &= -\cos(z) - \sin(z) \end{aligned}$$

6.11 Evaluate the following iterated integrals:

$$\int_2^4 \int_3^5 dydx \qquad \int_0^1 \int_0^1 x^{\frac{3}{2}} y^{\frac{2}{3}} dxdy$$

$$\int_1^3 \int_1^x \frac{x}{y} dydx \qquad \int_0^\pi \int_0^x y\cos(x) dydx$$

$$\int_0^1 \int_0^y (x+y^2) dxdy \qquad \int_1^2 \int_1^{\sqrt{2-y}} ydxdy$$

$$\int_0^1 \int_0^1 \int_0^1 \sqrt{1-x-y-z}\,dxdydz \qquad \int_2^e \int_e^3 \frac{1}{u}\frac{1}{v} dudv$$

$$\int_0^1 \int_0^{2-2x} \int_0^{x^2y^2} dzdydx \qquad \int_{-1}^1 \int_0^{1-y/3} xydxdy.$$

$$\int_2^4 \int_3^5 dydx = \int_2^4 y\Big|_3^5 dx = \int_2^4 2dx = 2x\Big|_2^4 = 4$$

$$\int_0^1 \int_0^1 x^{\frac{3}{2}} y^{\frac{2}{3}} dxdy = \int_0^1 y^{\frac{2}{3}} \frac{2}{5} x^{\frac{5}{2}} \Big|_0^1 dy$$

$$= \int_0^1 y^{\frac{2}{3}} \frac{2}{5} dy$$

$$= \left(\frac{2}{5}\right)\left(\frac{3}{5}\right) y^{\frac{5}{3}} \Big|_0^1 = \frac{6}{25}$$

$$\int_1^3 \int_1^x \frac{x}{y} dydx = \int_1^3 \int_1^x x \log(y) \Big|_1^x dx$$

$$= \int_1^3 [x \log(x) - x \log(1)]\, dx$$

$$(u = \log(x), dv = x, du = 1/x, v = \frac{1}{2}x^2)$$

$$= \log(x)\left(\frac{1}{2}\right)(x^2) - \frac{1}{4}(x^2) - \log(1)\left(\frac{1}{2}x^2\right)\Big|_1^3$$

$$= 2.943755$$

$$
\begin{aligned}
\int_0^\pi \int_0^x y\cos(x)dydx &= \int_0^\pi \frac{1}{2}y^2\cos(x)\Big|_{y=0}^{y=x} dx \\
&= \int_0^\pi \frac{1}{2}x^2\cos(x)dx \\
&\begin{pmatrix} u = \frac{1}{2}x^2 & dv = \cos(x) \\ du = x & v = -\sin(x) \end{pmatrix} \\
&= uv - \int vdu = (\frac{1}{2}x^2)(-\sin(x)) \\
&\qquad - \int_0^\pi (-\sin(x))xdx \\
&\begin{pmatrix} u = x & dv = \sin(x) \\ du = 1 & v = \cos(x) \end{pmatrix} \\
&= -\frac{1}{2}x^2\sin(x) + x\cos(x) - \int_0^\pi \cos(x)dx \\
&= -\frac{1}{2}x^2\sin(x) + x\cos(x) - \sin(x)\Big|_{x=0}^{x=\pi} \\
&= \pi
\end{aligned}
$$

Note: it helps to remember that $\frac{d}{dx}\sin(x) = \cos(x)$ implies $\int \frac{d}{dx}\sin(x) = \int\cos(x)$ meaning $\int\cos(x) = \sin(x)$, and $\frac{d}{dx}\cos(x) = -\sin(x)$ implies $\int \frac{d}{dx}\cos(x) = -\int\sin(x)$ meaning $\int\sin(x) = -\cos(x)$.

$$\int_0^1 \int_0^y (x+y^2)dxdy = \int_0^1 \left(\frac{1}{2}x^2 + xy^2\right)\bigg|_{x=0}^{x=1} dy$$
$$= \int_0^1 \left(\frac{1}{2}y^2 + y^3\right) dy$$
$$= \left(\frac{1}{6}y^3 + \frac{1}{4}y^4\right)\bigg|_0^1$$
$$= \frac{1}{6} + \frac{1}{4} = \frac{5}{12}$$

$$\int_1^2 \int_1^{\sqrt{2-y}} ydxdy = \int_1^2 [xy]\bigg|_{x=1}^{x=\sqrt{2-y}} dy$$
$$= \int_1^2 \left[(2-y)^{\frac{1}{2}} - y\right] dy$$
$$\begin{pmatrix} u = y & dv = (2-y)^{\frac{1}{2}} \\ du = 1 & v = \frac{2}{3}(2-y)^{\frac{3}{2}} \end{pmatrix}$$
$$= (y)\left(-\frac{2}{3}(2-y)^{\frac{3}{2}}\right)\bigg|_{y=1}^{y=2} - \int_1^2 \left(-\frac{2}{3}(2-y)^{\frac{3}{2}}(1)\right) dy - \frac{1}{2}y^2\bigg|_{y=1}^{y=2}$$
$$= \left[0 - \left(-\frac{2}{3}\right)\right] + \frac{2}{3}\left[-\frac{2}{5}(2-y)^5 2\right]\bigg|_{y=1}^{y=2} - \frac{1}{2}(2^2 - 1^2)$$
$$= \frac{2}{3} + \frac{4}{15} - \frac{3}{2} = -\frac{17}{30}$$

$$
\begin{aligned}
&\int_0^1 \int_0^1 \int_0^1 \sqrt{1-x-y-z}\,dxdydz \\
&= \int_0^1 \int_0^1 \left(-\frac{2}{3}(1-x-y-z)^{\frac{3}{2}}\right)\Big|_{x=0}^{x=1} \\
&= \int_0^1 \int_0^1 \left[-\frac{2}{3}(-y-z)^{\frac{3}{2}} + \frac{2}{3}(1-y-z)^{\frac{3}{2}}\right] dydz \\
&= \int_0^1 \left[-\left(\frac{2}{3}\right)\left(\frac{2}{5}\right)(-y-z)^{\frac{5}{2}}(-1)\right. \\
&\qquad\qquad \left. + \left(\frac{2}{3}\right)\left(\frac{2}{5}\right)(1-y-z)^{\frac{5}{2}}(-1)\right]\Big|_{y=0}^{y=1} dz \\
&= \frac{4}{15}\int_0^1 \left[(-y-z)^{\frac{5}{2}} - (1-y-z)^{\frac{5}{2}}\right]\Big|_{y=0}^{y=1} dz \\
&= \frac{4}{15}\int_0^1 \left[(-1-z)^{\frac{5}{2}} - (-z)^{\frac{5}{2}} - (-z)^{\frac{5}{2}} + (1-z)^{\frac{5}{2}}\right] dz \\
&= \frac{4}{15}\int_0^1 \left[(-1-z)^{\frac{5}{2}} + (1-z)^{\frac{5}{2}} - 2(-z)^{\frac{5}{2}}\right] dz \\
&= \frac{4}{15}\int_0^1 \left[\frac{2}{7}(-1-z)^{\frac{7}{2}}(-1) + \frac{2}{7}(1-z)^{\frac{7}{2}}(-1) - \frac{4}{7}(-z)^{\frac{7}{2}}(-1)\right]\Big|_{z=0}^{z=1} \\
&= \frac{4}{15}\frac{2}{7}\left[-(-1-z)^{\frac{7}{2}} - \frac{2}{7}(1-z)^{\frac{7}{2}} + 2(-z)^{\frac{7}{2}}\right]\Big|_{z=0}^{z=1} \\
&= \frac{8}{105}[-11.31371i - 0 + 2i + i + 1 - 0] \\
&= \frac{8}{105}[-8.31371i + 1]
\end{aligned}
$$

where $i = \sqrt{-1}$ (the “imaginary number”).

$$\int_2^e \int_e^3 \frac{1}{u}\frac{1}{v} du dv = \int_2^e \frac{1}{v} \log(u)\Big|_{u=e}^{u=3} dv$$

$$= \int_2^e \frac{1}{v}(\log(3) - \log(e)) dv$$

$$= (\log(3) - 1) \int_2^e \frac{1}{v} dv$$

$$= (\log(3) - 1)(\log(v))\Big|_{v=2}^{v=e}$$

$$= (\log(3) - 1)(1 - \log(2))$$

$$= \log(3) - \log(3)\log(2) - 1 + \log(2)$$

$$= 0.03025946$$

$$\int_0^1 \int_0^{2-2x} \int_0^{x^2y^2} dz dy dx = \int_0^1 \int_0^{2-2x} \left[x^2y^2\right] dy dx$$

$$= \int_0^1 \left[x^2 \left(\frac{1}{3}y^3\right)\right] \Big|_{y=0}^{y=2-2x} dx$$

$$= \frac{1}{3} \int_0^1 x^2(2-2x)^3 dx$$

$$= \frac{1}{3} \int_0^1 (4x^2 - 8x^3 + 4x^4) dx$$

$$= \frac{1}{3} \left[\frac{4}{3}x^3 - \frac{8}{4}x^4 + \frac{4}{5}x^5\right] \Big|_{x=0}^{x=1}$$

$$= \frac{1}{3}(\frac{4}{3} - 2 + \frac{4}{5}) = \frac{2}{45}$$

$$\int_{-1}^{1}\int_{0}^{1-y/3} xydxdy = \int_{-1}^{1}\left[\frac{1}{2}x^2y\right]\Big|_{x=0}^{x=1-y/3} dy$$

$$= \frac{1}{2}\int_{-1}^{1}(1-y/3)^2 ydy$$

$$= \frac{1}{2}\int_{-1}^{1}\left(1-\frac{2}{3}y+\frac{1}{9}y^2\right)dy$$

$$\frac{1}{2}\left[y-\frac{2}{3}\frac{1}{2}y^2+\frac{1}{9}\frac{1}{3}y^3\right]\Big|_{-1}^{1}$$

$$= \frac{1}{2}\left[\left(1-\frac{1}{3}+\frac{1}{27}\right)-(-1-\frac{1}{3}-\frac{1}{27})\right]$$

$$= \frac{28}{27}$$

6.13 Show whether the following series are convergent or divergent. If the series is convergent, find the sum.

$$\sum_{i=0}^{\infty}\frac{1}{3^i} \qquad \sum_{r=0}^{\infty}\frac{1}{(r+1)(r+2)} \qquad \sum_{i=1}^{\infty}\frac{1}{100i}$$

$$\sum_{k=1}^{\infty}k^{-\frac{1}{2}} \qquad \sum_{i=i}^{\infty}\left(\frac{i}{i-1}\right)^{\frac{1}{i}} \qquad \sum_{r=1}^{\infty}\sin\left(\frac{1}{r}\right)$$

$$\sum_{i=1}^{\infty}\frac{i^3+2i^2-i+3}{2i^5+3i-3} \qquad \sum_{r=1}^{\infty}\frac{2r+1}{(\log(r))^r} \qquad \sum_{i=1}^{\infty}\frac{2^i}{i^2}$$

For $\sum_{i=0}^{\infty}\frac{1}{3^i}$:

$$I_\infty = \int_0^\infty \frac{1}{3^x}dx = \int_0^\infty 3^{-x}dx = \int_0^\infty \exp\left[\log(3^{-x})\right]dx$$

$$= \int_0^\infty \exp\left[-x\log(3)\right]dx = \exp\left[-x\log(3)\right](-\log(3))\Big|_0^\infty$$

$$= \left(\frac{1}{3^x}\right)(-\log(3))\Big|_0^\infty = \log(3)$$

so it converges.

For $\sum_{r=0}^{\infty} \frac{1}{(r+1)(r+2)}$, we need to be somewhat creative. Consider first the first few values, continuing on:

$$S_\infty = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \ldots$$

which can be expressed as:

$$S_\infty = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{5}\right)$$
$$+ \left(\frac{1}{5} - \frac{1}{6}\right) + \ldots + \left(\frac{1}{r-1} - \frac{1}{r}\right) + \left(\frac{1}{r} - \frac{1}{r+1}\right) + \ldots$$

This then "telescopes" down to two terms (like Example 6.11) since consecutive values all cancel except the first and the last, producing:

$$S_\infty = 1 + \frac{1}{\infty} = 1$$

The series $\sum_{i=1}^{\infty} \frac{1}{100i}$ can be re-expressed as $\frac{1}{100} \sum_{i=1}^{\infty} \frac{1}{i}$, which is just a constant times a series that we know to be divergent. Therefore this is a divergent series.

The sum $\sum_{k=1}^{\infty} k^{-\frac{1}{2}}$ fits into the following principle. A sum defined by $\sum_{k=1}^{\infty} \frac{1}{k^\rho}$ converges for $\rho > 1$ and diverges otherwise. Therefore this is a divergent sum. We can also, less generally, note that:

$$\int_1^\infty x^{-\frac{1}{2}} dx = \frac{1}{2} x^{\frac{1}{2}} \Big|_1^\infty = \infty$$

For the sum $\sum_{i=2}^{\infty} \left(\frac{i}{i-1}\right)^{\frac{1}{i}}$, we know $\lim\limits_{n\to\infty} \left(\frac{1}{1-1/n}\right)^{\frac{1}{n}}$ never produces values lower than one, so this series fails the first integral test on page 260.

The series $\sum_{r=1}^{\infty} \sin\left(\frac{1}{r}\right)$ is convergent since:

$$\lim_{n\to\infty} \sin(1/n) = 0$$

For $\sum_{i=1}^{\infty} \frac{i^3+2i^2-i+3}{2i^5+3i-3}$, look at the limit:

$$\lim_{n\to\infty} \frac{n^3+2n^2-n+3}{2n^5+3n-3} = \lim_{n\to\infty} \frac{1/n^2+2/n^3-1/n^4+3/n^5}{2+3/n^4-3/n^5} = 0$$

which shows that the series converges.

For $\sum_{r=1}^{\infty} \frac{2r+1}{(\log(r))^r}$ use the ratio test:

$$\lim_{n\to\infty} \left[\frac{\frac{2(n+1)+1}{(\log(n+1))^n}}{\frac{2(n)+1}{(\log(n))^n}} \right]$$

$$= \lim_{n\to\infty} \frac{(2n+3)(\log(n))^n}{(2n+1)(\log(n+1))^{n+1}}$$

$$= \lim_{n\to\infty} \frac{(2n+3)(\log(n))^n}{(2n+1)(\log(n+1))^n \log(n+1)}$$

$$= \lim_{n\to\infty} \left(\frac{2n+3}{2n+1}\right)\left(\frac{1}{\log(n+1)}\right)\left(\frac{\log(n)}{\log(n+1)}\right)$$

$$= \lim_{n\to\infty} \left(\left(\frac{2n+1}{2n+1}\right) + \left(\frac{2}{2n+1}\right)\right)\left(\frac{1}{\log(n+1)}\right)\left(\frac{\log(n)}{\log(n+1)}\right)$$

$$= (1+0)(0)(1) = 0$$

So the series converges.

Using the ratio test for $\sum_{i=1}^{\infty} \frac{2^i}{i^2}$,

$$\lim_{n\to\infty} \frac{2^n/n^2}{2^{n+1}/(n+1)^2} = \lim_{n\to\infty} \frac{1}{2} \frac{n^2}{(n+1)^2} = \frac{1}{2}$$

6.15 Write the following repeating decimals forms in series notation:

$0.3333\ldots$	$0.43114311\ldots$	$0.484848\ldots$
$0.1234512345\ldots$	$555551555551\ldots$	$0.221222223221222223\ldots.$

$$\begin{aligned} 0.3333\ldots &= \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \ldots \\ &= \sum_{k=1}^{\infty} \frac{3}{10^k} \end{aligned}$$

$$\begin{aligned} 0.43114311\ldots &= \frac{4311}{10000} + \frac{4311}{100000000} + \ldots \\ &= \sum_{k=1}^{\infty} \frac{4311}{10^{4k}} \end{aligned}$$

$$\begin{aligned} 0.484848\ldots &= \frac{48}{100} + \frac{48}{10000} + \frac{48}{100000} + \ldots \\ &= \sum_{k=1}^{\infty} \frac{48}{10^{2k}} \end{aligned}$$

$$\begin{aligned} 0.1234512345\ldots &= \frac{12345}{100000} + \frac{12345}{10000000000} + \ldots \\ &= \sum_{k=1}^{\infty} \frac{48}{10^{5k}} \end{aligned}$$

$$\begin{aligned} 0.555551555551\ldots &= \frac{55551}{1000000} + \frac{555551}{1000000000000} + \ldots \\ &= \sum_{k=1}^{\infty} \frac{555551}{10^{6k}} \end{aligned}$$

$$\begin{aligned} 0.221222223221222223\ldots &= \frac{22}{1000000} + \frac{555551}{1000000000000} + \ldots \\ &= \sum_{k=1}^{\infty} \sum_{\ell=1}^{3} \frac{220+\ell}{10^{3k}} \end{aligned}$$

6.17 Find the Maclaurin series for $\sin(x)$ and $\cos(x)$. What do you observe?

$$\sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} = \frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots$$

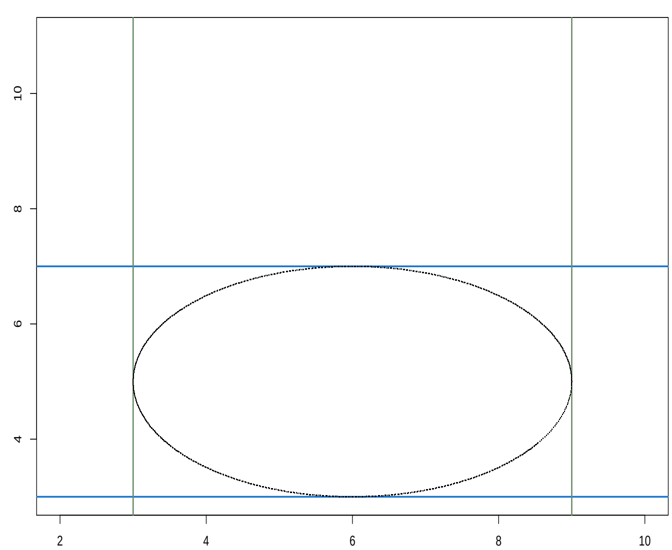
$$\cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} = \frac{x^0}{0!} - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots$$

The $2k$ versus $2k+1$ difference highlights the oscillatory difference in the two functions.

6.19 Find the vertical and horizontal tangent lines to the ellipse defined by

$$x = 6 + 3\cos(t) \qquad y = 5 + 2\sin(t).$$

This is an ellipse centered at $x = 6$, $y = 5$. We know that the derivative gives the tangent line so $f^{'}(x, y) = (-3\sin(t), 2\cos(t))$ is the function we need. The horizontal tangent lines are where $x = 6$ and $\cos(t) = 0$ since the slope is flat. This occurs at $t = \pi/2$ and $t = 3\pi/2$. Therefore the lines are defined by $y = 5+2\sin(\pi/2)$ and $y = 5+2*\sin(\pi/2)$. These are indicated in the figure below with blue lines. The horizontal tangent lines exist where $y = 5$ on the function. If $y = 5$ is on the function, then $5 = 5 + 2\sin(t)$. Since $\sin(t) = 0$ occurs at π and 2π, we can use $x = 6 + 3\cos(\pi)$ and $x = 6 + 3\cos(2\pi)$. These lines are graphed below in green.



6.21 Given $\mathbf{f}(t) = \frac{1}{t}\mathbf{i} + \frac{1}{t^3}$, find the first three orders of derivatives. Solve for $t = 2$.

$$\mathbf{f}^{'}(t) = -t^{-2}\mathbf{i} - 3t^{-4}$$

$$\mathbf{f}^{'}(t = 2) = -\frac{1}{4}\mathbf{i} - \frac{3}{16}$$

$$\mathbf{f}^{''}(t) = 2t^{-3}\mathbf{i} + 12t^{-5}$$

$$\mathbf{f}^{''}(t = 2) = \frac{1}{4}\mathbf{i} + \frac{3}{8}$$

$$\mathbf{f}^{'''}(t) = -6t^{-4}\mathbf{i} + 60t^{-6}$$

$$\mathbf{f}^{'''}(t = 2) = -\frac{3}{8}\mathbf{i} - \frac{15}{16}$$

6.23 A number of seemingly counterintuitive voting principles can actually be proven mathematically. For instance, Brams and O'Leary (1970) claimed that "If three kinds of votes are allowed in a voting body, the probability that two randomly selected members disagree on a roll call will be maximized when one-third of the members vote 'yes,' one-third 'no,' and one third 'abstain."' The proof of this statement rests on the premise that their probability of disagreement function is maximized when $y = n = a = t/3$, where y is the number voting yes, n is the number voting no, a is the number abstaining, and these are assumed to divide equally into the total number of voters t. The disagreement function is given by

$$p(DG) = \frac{2(yn + ya + na)}{(y + n + a)(y + n + a - 1)}.$$

Use the Lagrange multiplier method to demonstrate their claim by first taking four partial derivatives of $p(DG) - \lambda(t - y - n - a)$ with respect to y, n,

a, λ (the Lagrange multiplier); setting these equations equal to zero; and solving the four equations for the four unknowns.

This is a long and tedious calculation. Assign it only if you are irritated with your class. There is certainly a linear algebra short-cut that can be developed, but the most obvious way is brute-force calculus and algebra. I will provide only one of the four processes here and omit the final (and lengthy) step of solving. This last step should be done with matrices.

$$\begin{aligned}
&\frac{\partial}{\partial y} L(DG, \lambda) \\
&= 2(y+a)(y+n+a)^{-1}(y+n+a-1)^{-1} \\
&\quad + 2(yn+ya+na)(-1)(y+n+a)^{-2}(1)(y+n+a-1)^{-1} \\
&\quad + 2(yn+ya+na)(y+n+a)^{-1}(-1)(y+n+a-1)^{-2}(1) + \lambda \\
&= \frac{2(y+a)}{(y+n+a)(y+n+a-1)} - \frac{2(yn+ya+na)}{(y+n+a)^2(y+n+a-1)} \\
&\quad - \frac{2(yn+ya+na)}{(y+n+a)(y+n+a-a)^2} + \lambda \equiv 0
\end{aligned}$$

This begins a lengthy algebraic cleanup. . .

$$
\begin{aligned}
0 &= (y+a)(y+n+a)(y+n+a-1) \\
&\quad -(yn+ya+na)(y+n+a-1)-(yn+ya+na)(y+n+a) \\
&\quad +\frac{\lambda}{2}(y+n+a)^2(y+n+a-1)^2 \\
&= (y^2+yn+ya+ay+an+a^2)(y+n+a-1) \\
&\quad -(y^2n-yn^2+yna-yn+y^2a+yan \\
&\quad +ya^2-ya+nay+n^2a+na^2-na) \\
&\quad -(y^2n+yn^2+yna+y^2a+yan+ya^2) \\
&\quad +\frac{\lambda}{2}(y^2+yn+ya+ny+n^2+na+ay+an+a^2) \\
&\quad \times(y^2+yn+ya-y-y+ny+n^2+na-n+ay \\
&\quad +an+a^2-a-y-n-a+1) \\
&= (y^2+yn+2ay+an+a^2)(y+n+a-1) \\
&\quad -y^2n-y^2a-yn^2-n^2a-ya^2-na^2-2yna \\
&\quad +yn+ya+na-y^2n-yn^2-y^2a-2yna-ya^2 \\
&\quad +\frac{\lambda}{2}(y^2+2ny+2ay+2an+n^2+a^2) \\
&\times(y^2+n^2+a^2+2ay+2ny+2an-2y-2n-2a+1)
\end{aligned}
$$

$$
\begin{aligned}
&= y^3 + y^2n + y^2a - y^2 + y^2n + yn^2 + yna - yn + 2y^2a \\
&\quad + 2ayn + 2a^2y - 2ay + any + an^2 + a^2n - an + a^2y \\
&\quad + a^2n + a^3 - a^2 - 2y^2n - 2y^2a - 2ya^2 - 2yn^2 - n^2a \\
&\quad - na^2 - 4yna + yn + ya + na \\
&\quad + \frac{\lambda}{2}(y^4 + y^2n^2 + y^2a^2 + 2ay^3 + 2ny^3 + 2any^2 - 2y^3 \\
&\quad - 2ny^2 - 2ay^2 + y^2 + 2ny^3 + 2n^3y + 2nya^2 + 4nay^2 \\
&\quad + 4n^2y^2 + 4n^2y^2 + 2an^2y - 4ny^2 - 4n^2y - 4any + 2ny \\
&\quad + 2ay^3 + 2ayn^2 + 2a^3y - 4a^2y^2 + 4any^2 + 4a^2yn \\
&\quad - 4ay^2 - 4any - 4a^2y + 2ay + 2any^2 + 2an^3 + 2a^3n \\
&\quad + 4a^2ny + 4an^2y + 4a^2n^2 - 4any - 4an^2 - 4a^2n + 2an \\
&\quad + n^2y^2 + n^4 + n^2a^2 + 2ayn^2 + 2n^3y + 2an^3 - 2yn^2 \\
&\quad - 2n^3 - 2an^2 + n^2 + a^2y^2 + a^2n^2 + a^4 + 2a^3y \\
&\quad + 2a^2ny + 2a^3n + 2an^3 - 2yn^2 - 2n^3 - 2an^2 + n^2)
\end{aligned}
$$

$$
\begin{aligned}
&= y^3 - y^2 - yn^2 + y^2a + a^2n + a^3 - a^2 + a^2y - ay \\
&\quad + \frac{\lambda}{2}(y^4 + 4ay^3 + 4n^3y + 4an^3 - 2n^3 + 4a^3y + 2an^3 + a^4 \\
&\quad + 4ny^3 + 5n^2y^2 + 2n^2a^2 - 3a^2y^2 + 2a^3n + 4a^2n^2 - 2y^3 \\
&\quad - 6ny^2 - 4yn^2 + 2a^3n + n^4 - 2ay^2 - 4n^2y - 4a^2y - 4a^2n \\
&\quad - 2n^3 - 8an^2 + y^2 + 2n^2 + 12nya^2 + 10an^2y + 12nay^2 \\
&\quad + y^2n^2 + y^2a^2 - 12any + 2ny - 4ay^2 + 2ay + 2an)
\end{aligned}
$$

7

Probability Theory

7.1 A fair coin is tossed 20 times and produces 20 heads. What is the probability that it will give a tails on the 21st try?

It is still 0.5.

7.3 Prove that $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$.

$$\frac{(n-1)!}{(n-k-1)!k!} + \frac{(n-1)!}{(n-k)!(k-1)!}$$

$$\frac{(n-k)(n-1)! + k(n-1)!}{(n-k)!k!}$$

$$\frac{(n-1)!(n-k+k)}{(n-k)!k!}$$

7.5 Suppose you had a pair of four-sided dice (they exist), so the set of possible outcomes from summing the results from a single toss is $\{2, 3, 4, 5, 6, 7, 8\}$. Determine the probability of each of these outcomes.

$$\#2 outcomes = (1,1) \qquad p(2) = 1/16$$

$$\#3 outcomes = (2,1),(1,2) \qquad p(3) = 2/16$$

$$\#4 outcomes = (3,1),(1,3),(2,2) \qquad p(4) = 3/16$$

$$\#5 outcomes = (4,1),(1,4),(2,3),(3,2) \qquad p(5) = 4/16$$

$$\#6 outcomes = (4,2),(2,4),(3,3) \qquad p(6) = 3/16$$

$$\#7 outcomes = (4,3),(3,4) \qquad p(7) = 2/16$$

$$\#8 outcomes = (4,4) \qquad p(8) = 1/16$$

7.7 Prove de Morgan's Laws for two sets A and B.

Show $(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$.

Suppose: $X \in (A \cup B)^{\complement}$.

Then:$X \notin A$ and $X \notin B$.

So:$X \in A^{\complement} \cap B^{\complement}$.

Show $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$.

Suppose: $X \in (A \cap B)^{\complement}$.

Then: X is in one of:$(A \setminus B), (B \setminus A), (A^{\complement} \cap B^{\complement})$.

A partition of the sample space consists of these three sets along with $A \cap B$.So:$X \in (A \cap B)^{\complement}$.

7.9 If $(D|H) = 0.5,\ p(D) = 1$, and $p(H) = 0.1$, what is the probability that H is true given D?

We know:

$$p(H|D) = p(H,D)/p(D)$$

and

$$p(D|H) = p(D, H)/p(H)$$

so

$$p(H|D)p(D) = p(D|H)p(H)$$

and therefore $p(H|D) = p(D|H)p(H)/p(D) = 0.05$.

7.11 Show that the Theorem of Total Probability also works when either of the two sets is the null set.

$p(A) = P(A \cap B) + P(A \cap B^c)$. Set $A = \varnothing$. So $p(\varnothing) = p(\varnothing \cap B) + p(\varnothing \cap B^c) = p(\varnothing) + p(\varnothing) = 0$. Set $B = \varnothing$. So $p(A) = p(A \cap \varnothing) + p(A \cap \varnothing^c)$, $\varnothing^c = \Omega$, $p(A) = p(\varnothing) + p(A) = p(A)$.

7.13 If events A and B are independent, prove that A^{C} and B^{C} are also independent. Can you say that A and A^{C} are independent? Show your logic.

If A and B are independent, then:

$$\begin{aligned} p(A \cap B) &= p(A)p(B) \\ &= [1 - p(A^{\mathsf{C}})][1 - p(B^{\mathsf{C}})] \\ &= 1 - p(B^{\mathsf{C}}) - p(A^{\mathsf{C}}) + p(A^{\mathsf{C}})p(B^{\mathsf{C}}). \end{aligned}$$

By the law of complementation and de Morgan's Law:

$$\begin{aligned} p(A \cap B) &= 1 - p(A \cap B)^{\mathsf{C}} \\ &= 1 - p(A^{\mathsf{C}}) \cup p(B^{\mathsf{C}}) \\ &= 1 - \left[p(A^{\mathsf{C}}) + p(B^{\mathsf{C}}) - p(A^{\mathsf{C}} \cap B^{\mathsf{C}})\right]. \end{aligned}$$

Now put these two together and simplify:

$$1 - p(B^{\mathsf{C}}) - p(A^{\mathsf{C}}) + p(A^{\mathsf{C}})p(B^{\mathsf{C}})$$

$$= 1 - \left[p(A^{\mathsf{C}}) + p(B^{\mathsf{C}}) - p(A^{\mathsf{C}} \cap B^{\mathsf{C}})\right]$$

$$p(A^{\mathsf{C}})p(B^{\mathsf{C}}) = p(A^{\mathsf{C}} \cap B^{\mathsf{C}}).$$

For the second part, if A and $A^{\complement}$ are independent then:

$$P(A \cap A^{\complement}) = P(A)P(A^{\complement}).$$

We know that $P(A \cap A^{\complement}) = 0$ by definition of complementation. Since $P(A^{\complement}) = 1 - P(A)$, then:

$$P(A)P(A^{\complement}) = P(A)(1 - P(A))$$

which can be zero only for the pathological case where $P(A) = 0$ then these are not independent.

7.15 Use this joint probability distribution

		X 0	1	2
	0	0.10	0.10	0.01
Y	1	0.02	0.10	0.20
	2	0.30	0.10	0.07

to compute the following:

a) $p(X < 2)$.
b) $p(X < 2|Y < 2)$.
c) $p(Y = 2|X \leq 1)$.
d) $p(X = 1|Y = 1)$.
e) $p(Y > 0|X > 0)$.

a) $p(X < 2) = 0.42 + 0.3 = 0.72$.
b) $p(X < 2|Y < 2) = \frac{p(X<2,Y<2)}{p(Y<2)} = \frac{0.32}{0.21+0.32} = 0.6038$.
c) $p(Y = 2|X \leq 1) = \frac{p(Y=2,X\leq 1)}{p(X\leq 1)} = \frac{0.3+0.1}{0.72} = 0.5555$.
d) $p(X = 1|Y = 1) = \frac{0.10}{0.02+0.10+0.20} = 0.3125$.
e) $p(Y > 0|X > 0) = \frac{0.10+0.20+0.10+0.07}{0.10+0.20+0.10+0.07+0.10+0.01} = 0.7015$.

7.17 Someone claims they can identify four different brands of beer by taste. An experiment is set up (off campus of course) to test her ability in which she is given each of the four beers one at a time without labels or any visual identification.

a) How many different ways can the four beers be presented to her one at a time?

b) What is the probability that she will correctly identify all four brands simply by guessing?

c) What is the probability that she will incorrectly identify only one beer simply by guessing (assume she does not duplicate an answer)?

d) Is the event that she correctly identifies the second beer disjoint with the event that she incorrectly identifies the fourth beer?

a) Call the four beers: A,B,C,D. So if A is given first and B is given second there are 2 ways to give C and D. If A is given first, then there are 6 ways to give B, C, and D (try this). As it turns out there are 4! =4*3*2*1 =24 total permutations. You could also just have listed them all.

b) There are 24 permutations and only 1 way for her to get them all right so the probability is $\frac{1}{24}$.

c) She cannot incorrectly identify 1 beer since there would be no other beer to get the name she missed. The probability is therefore zero.

d) Since she can correctly identify the second *and* incorrectly identify the fourth in the same test, the intersection is not the null set. Therefore the two events are not disjoint.

7.19 Suppose your professor of political theory put 17 books on reserve in the library. Of these, 9 were written by Greek philosophers and the rest were written by German philosophers. You have already read all of the Greeks, but none of the Germans, and you have to ask for the books one at a time. Assuming you left the syllabus at home, and you have to ask for the books at random (equally likely) by call letters:

a) What is the probability that you have to ask for at least three books before getting a German philosopher?

b) What is the highest possible number of times you would have to ask for a book before receiving a German philosopher?

a) Let X equal the number of times you have to ask for a book:
$p(X \geq 3) = 1 - p(X < 3) = 1 - p(X = 1) - p(X = 2)$
$p(X = 1) = p(\text{German on first try}) = \frac{8}{17} = 0.4706$
$p(X = 2) = p(\text{Greek on first try})p(\text{German on second try})$
$= \frac{9}{17}\frac{8}{16} = 0.2647$
$p(X \geq 3) = 1.0 - 0.4706 - 0.2647 = 0.2648$

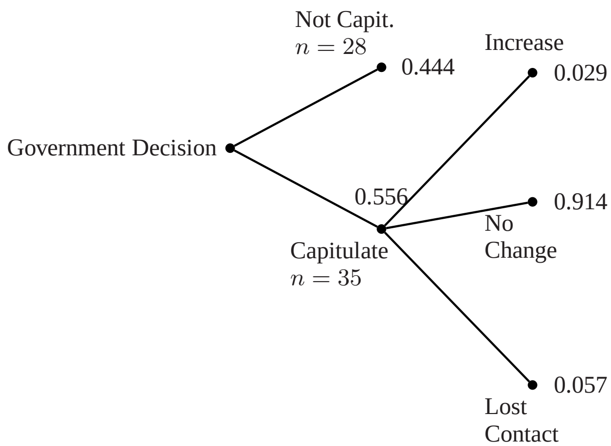
b) Since there are 9 Greek philosophy books, even if you were extremely unlucky and got these 9 books in the first 9 tries, you are guaranteed to get a German philosophy book on the 10^{th} try.

7.21 In a given town, 40% of the voters are Democrats and 60% are Republican. The president's budget is supported by 50% of the Democrats and 90% of the Republicans. If a randomly (equally likely) selected voter is found to support the president's budget, what is the probability that they are a Democrat?

$p(D|S) = p(D)p(S|D)/p(S) = p(D)p(S|D)/(p(D)p(S|D)+p(R)p(S|R)) = (0.4) * (0.5)/((0.4) * (0.5) + (0.6) * (0.9)) = 0.27.$

7.23 Assume that 2% of the population of the United States are members of some extremist militia group, ($p(M) = 0.02$), a fact that some members might not readily admit to an interviewer. We develop a survey that is 95% accurate on positive classification, $p(C|M) = 0.95$, and 97% accurate on negative classification, $p(C^{\complement}|M^{\complement}) = 0.97$. Using Bayes' Law, derive the probability that someone positively classified by the survey as being a militia member really is a militia member. (Hint: Draw a Venn diagram to get $p(C)$ and think about the Theorem of Total Probability).

The unconditional probability of classifying a respondent as a militia member results from accumulation of the probability across the sample space of survey events using the Total Probability Theorem: $p(C) = p(C \cap M) +$



$p(C \cap M^{\complement}) = p(C|M)p(M) + [1 - p(C^{\complement}|M^{\complement})]p(M^{\complement}) = (0.95)(0.02) + (0.03)(0.98) \cong 0.05$. This is then used in Bayes Law: $p(M|C) = \frac{p(M)}{p(C)}p(C|M) = \frac{0.02}{0.05}(0.95) = 0.38$.

7.25 Corsi (1981) examined political terrorism, responses to terrorist acts, and the counter-response of the terrorists for 1970 to 1974. For the type of events where a target is seized and held at an unknown site (like kidnapping) he found that 55.6% ($n = 35$) of the time the government involved capitulated. Given that this happened, 2.9% of the time the terrorists increased their demands, 91.4% of the time there was no change in these demands, and 5.7% of the time contact is lost. Of these three events, the number of times that there was known to be no damage or death was 1, 31, and 1, respectively. Construct a tree diagram that contains the conditional probabilities at each level.

The tree diagram is given below. The counts at the end are interesting but not part of the problem.

8

Random Variables

8.1 Indicate the level of measurement and which measure(s) of central tendency can be used for the following:

a) college education: none, AA, BA/BS, JD, MD/DVM/DDO, Ph.D.;
b) letter grades;
c) income given as 0–10K, 10–20K, 30–50K, 50–80K, 100K+;
d) distance of commute from home to work;
e) marital status: single, married, widowed, divorced;
f) working status: employed, unemployed, retired, student;
g) governmental level: local, state, federal, international;
h) party: Democrat, Republican, Green, Bull Moose.

a) Nominal.
b) Ordinal.
c) Ordinal.
d) Ratio.
e) Nominal.
f) Nominal.
g) Ordinal.
h) Nominal.

8.3 Morrison (1977) gave the following data for Supreme Court vacancies from 1837 to 1932:

Number of Vacancies/Year	0	1	2	3	4+
Number of Years for Event	59	27	9	1	0

Fit a distribution to these data, estimating any necessary parameters. Using this model, construct a table of expected versus observed frequencies by year.

The obvious solution is to fit a Poisson model by estimating λ from these data. This produces: $mean = 0.5$, $var = 0.5053$ from the data. Note that this implies a good fit for the Poisson since the mean and variance are nearly identical. Predictions from this model are:

observed	59	27	9	1	0
expected	58.226	29.113	7.278	1.213	0.170

which are produced from plugging in the observed vacancies as x information.

8.5 Suppose you had a Poisson process with intensity parameter $\lambda = 5$. What is the probability of getting exactly 7 events? What is the probability of getting exactly 3 events? These values are the same distance from the expected value of the Poisson distribution, so why are they different?

The probability of getting 5 events:

$$p(7|\lambda = 5) = \frac{e^{-5}5^7}{7!} = 0.1044449.$$

The probability of getting 3 events:

$$p(3|\lambda = 5) = \frac{e^{-5}5^3}{3!} = 0.1403739.$$

These are different because, unlike the normal, the shape of the Poisson distribution is asymmetrical around it's mean.

8.7 Let X be the event that a single die is rolled and the resulting number is even. Let Y be the event describing the actual number that results from the roll (1–6). Prove the independence or nonindependence of these two events.

$$p(even)p(1) = \frac{3}{6}\frac{1}{6} = \frac{1}{12}, \qquad p(even, 1) = 0$$

$$p(even)p(2) = \frac{3}{6}\frac{1}{6} = \frac{1}{12}, \qquad p(even, 2) = \frac{1}{6}$$

$$p(even)p(3) = \frac{3}{6}\frac{1}{6} = \frac{1}{12}, \qquad p(even, 3) = 0$$

$$p(even)p(4) = \frac{3}{6}\frac{1}{6} = \frac{1}{12}, \qquad p(even, 4) = \frac{1}{6}$$

$$p(even)p(5) = \frac{3}{6}\frac{1}{6} = \frac{1}{12}, \qquad p(even, 5) = 0$$

$$p(even)p(6) = \frac{3}{6}\frac{1}{6} = \frac{1}{12}, \qquad p(even, 6) = \frac{1}{6}.$$

So the events are clearly not independent.

8.9 Suppose we have a PMF with the following characteristics: $p(X = -2) = \frac{1}{5}$, $p(X = -1) = \frac{1}{6}$, $p(X = 0) = \frac{1}{5}$, $p(X = 1) = \frac{1}{15}$, and $p(X = 2) = \frac{11}{30}$. Define the random variable $Y = X^2$. Derive the PMF of Y and prove that it is a PMF. Calculate the expected value and variance of Y.

This problem is greatly simplified because it there is a discrete sample space with a small number of outcomes. First, we specify the probabilities for Y:

$$p(Y = 0) = p(X = 0) = \frac{1}{5} = \frac{6}{30}$$

$$p(Y = 1) = p(X = -1) + p(X = 1) = \frac{1}{6} + \frac{1}{15} = \frac{7}{30}$$

$$p(Y = 4) = p(X = -2) + p(X = 2) = \frac{1}{5} + \frac{11}{30} = \frac{17}{30}.$$

To prove that this is a PMF, we need to show that it satisfies the Kolmogorov axioms:

- The probability of any realizable event is between zero and one:

$$0 < \frac{6}{30}, \frac{7}{30}, \frac{17}{30} < 1.$$

- Something happens with probability one:

$$\sum_{k=0}^{2} p(Y = k) = 1.$$

- The probability of unions of n pairwise disjoint events is the sum of their individual probabilities:

$$\begin{aligned} p(Y = 0 \cup 1) &= p(X = -1 \cup 0 \cup 1) \\ &= \frac{1}{6} + \frac{1}{5} + \frac{1}{15} = \frac{13}{30} \\ p(Y = 0) + p(Y = 1) &= \frac{6}{30} + \frac{7}{30} = \frac{13}{30} \\ p(Y = 1 \cup 4) &= P(X = -2 \cup -1 \cup 1 \cup 2) \\ &= \frac{1}{5} + \frac{1}{6} + \frac{1}{15} + \frac{11}{30} = \frac{24}{30} \\ P(Y = 1) + P(Y = 4) &= \frac{7}{30} + \frac{17}{30} = \frac{24}{30} \\ p(Y = 4 \cup 0) &= p(X = -2 \cup 0 \cup 2) \\ &= \frac{1}{5} + \frac{1}{5} + \frac{11}{30} = \frac{23}{30} \\ p(Y = 4) + p(Y = 0) &= \frac{17}{30} + \frac{6}{30} = \frac{23}{30}. \end{aligned}$$

The expected value of Y is:

$$E[Y] = 0\left(\frac{6}{30}\right) + 1\left(\frac{7}{30}\right) 4\left(\frac{17}{30}\right) = 2.5.$$

The expected value of Y^2 is:

$$E[Y] = 0^2\left(\frac{6}{30}\right) + 1^2\left(\frac{7}{30}\right) 4^2\left(\frac{17}{30}\right) = 9.3.$$

So the variance is:

$$\text{Var}[Y] = E[Y^2] - E[Y]^2 = 9.3 - 2.5^2 = 3.05.$$

8.11 Twenty developing countries each have a probability of military coup of 0.01 in any given year. We study these countries over a 10-year period.

a) How many coups do you expect in total?

b) What is the probability of four coups?

c) What is the probability that there will be no coups during this period?

Assuming independence across time and countries...

a) Expected total:

$$E[total] = \sum_{i=1}^{20}\sum_{j=1}^{10}(0.01) = 2$$

b) The probability of four coups:

$$p(T = 4) = p(4\,coups) + p(96\,non - coups)$$

$$= (1/100)^4 + (99/100)^{96} = 0.3810471$$

c) The probability no coups:

$$p(T = 0) = p(100\,non - coups) = (99/100)^{200} = 0.1339797$$

8.13 Use the exponential PDF to answer the following questions.

a) Prove that the exponential form *is* a PDF.

b) Derive the CDF.

c) Prove that the exponential distribution is a special case of the gamma distribution.

The exponential PDF is given by:

$$f(x|\beta) = \frac{1}{\beta}\exp[-x/\beta], \qquad 0 \le x < \infty, \quad 0 < \beta,$$

a) Showing that the exponential PDF satisfies the Kolmogorov axioms:

- The probability of any realizable event is between zero and one:

$$\lim_{x \to \infty} f(x|\beta) = 0$$

$$\lim_{x \to 0} f(x|\beta) = \frac{1}{\beta}$$

- Something happens with probability one:

$$\int_0^\infty \frac{1}{\beta} \exp[-x/\beta] dx = -\exp[-x/\beta] \Big|_1^\infty = 0 - (-1) = 1$$

- The probability of unions of n pairwise disjoint events is the sum of their individual probabilities: we know that for any $0 < a < b < c < \infty$,

$$p(a < X < b) + p(b < X < c])$$

$$= \int_a^b \frac{1}{\beta} \exp[-x/\beta] dx + \int_b^c \frac{1}{\beta} \exp[-x/\beta] dx$$

$$= \left(-\exp[-x/\beta] \Big|_a^b \right) + \left(-\exp[-x/\beta] \Big|_a^b \right)$$

$$= -\exp[-a/\beta] + \exp[-b/\beta] - \exp[-b/\beta] + \exp[-c/\beta]$$

$$= -\exp[-a/\beta] + \exp[-c/\beta]$$

$$= -\exp[-x/\beta] \Big|_a^c$$

$$= \int_a^c \frac{1}{\beta} \exp[-x/\beta] dx$$

$$= p(a < X < c)$$

b) The CDF is found by integrating from zero to x:

$$\int_0^x f(t|\beta)dt = -\exp[-t/\beta]\Big|_0^x$$

$$= 1 - \exp[-x/\beta]$$

where CDF values for $x < 0$ are 0.

c) The gamma distribution is given by:

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp[-x/\beta],$$

for $0 \le x < \infty, \quad 0 < \alpha, \beta$. Now set $\alpha = 1$ to establish a single special case:

$$f(x|\alpha = 1,\beta) = \frac{1}{\Gamma(1)\beta^{(1)}} x^{(1)-1} \exp[-x/\beta]$$

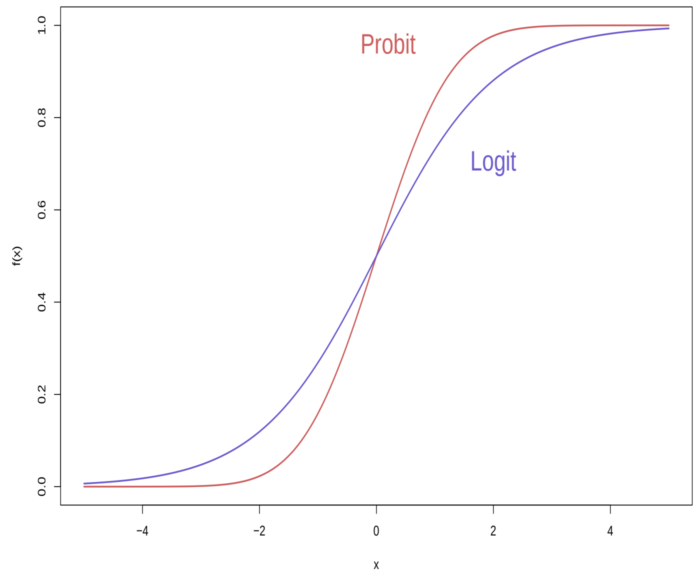
$$= \frac{1}{\beta}\exp[-x/\beta]$$

which of course is the exponential form above.

8.15 A function that can be used instead of the probit function is the logit function: $\Lambda(X) = \frac{\exp(X)}{1+\exp(X)}$. Plot both the logit function and the probit function in the same graph and compare. What differences do you observe?

This figure shows that the logit function has slightly "heavier" tails in that it asymptotes slower than the probit function. Note, though, that they intersect at $[0, 0.5]$. For the probit, this is a property of the standard normal CDF accounting for half the density at zero. For the logit function, it comes from $\exp(0) = 1$.

8.17 Prove that $E[Y|Y] = Y$.



This is easiest to show for the discrete case:

$$
\begin{aligned}
E[Y|Y] &= \sum_{i=1}^{k} Y_i p(Y_i|Y) \\
&= \sum (Y_i(0) \; \forall \; Y_i \neq Y) + Y_i(1) \text{ for } Y_i = Y \\
&= Y_i \ni (Y_i = Y) \\
&= Y
\end{aligned}
$$

8.19 Calculate the entropy of the distribution $\mathcal{B}(n = 5, p = 0.1)$ and the distribution $\mathcal{B}(n = 3, p = 0.5)$. Which one is greater? Why?

Use the natural log since comparing distributions makes this decision arbitrary:

$$H_1 = \sum_{i=1}^{5} p(X = i|n = 5, p = 0.1) \log(p(X = i|n = 5, p = 0.1))$$

$$= (0.59049)(-0.5268026) + (0.32805)(-1.1145892)$$

$$+ (0.07290)(-2.6186666) + (0.00810)(-4.8158912)$$

$$+ (0.00045)(-7.7062630) + (0.00001)(-11.5129255)$$

$$= -0.9102051$$

$$H_2 = \sum_{i=1}^{3} p(X = i|n = 3, p = 0.5) \log(p(X = i|n = 3, p = 0.5))$$

$$= (0.729)(-0.3160815) + (0.243)(-1.4146938)$$

$$+ (0.027)(-3.6119184) + (0.001)(-6.9077553)$$

$$= -0.6786236$$

So the second one has higher entropy even though there are fewer possible events. The reason is that the probability of $p = 0.5$ for the binomial produces the greatest possible variance.

8.21 Show that the second moment of the Cauchy distribution is infinite and therefore undefined.

We want $E[X^2]$ where X is distributed Cauchy. Start with the definition of variance for a random variable:

$$\text{Var}[X] = E[X^2] - E[X]^2$$

and rearrange for the quantity of interest:

$$E[X^2] = Var[X] + E[X]^2.$$

We know that for the Cauchy $E[X] = \infty$ (shown on page 382). So $E[X]^2$ must also be infinity as well as $Var[X]+E[X]^2$. Therefore the first moment property "pollutes" all higher moments by induction.

8.23 The following is a stem and leaf plot for 20 different observations (stem = tens digit). Use these data to answer the questions.

0	7	8	9	9	9	
1	0	1	5	7	7	9
2	0	1	1	8		
3	2	4				
4						
5						
6						
7	5					
8	3	9				

a) Is the median bigger or smaller than the mean?

b) Calculate the 10% trimmed mean.

c) Make a frequency distribution with relative *and* relative cumulative frequencies.

d) Calculate the standard deviation.

e) Identify the IQR.

a) The data are right-skewed so the median *must* be smaller than the mean: $\bar{X} = 26.7$, and $M_x = 18$.

b) $\bar{X}_{.10} = 21.6875$.

c) Tabulating:

0-9	5	0.25	0.25
Category	Freq.	Relative Freq.	Cumulative Freq.
0-9	5	0.25	0.25
10-19	6	0.30	0.55
20-29	4	0.20	0.75
30-39	2	0.10	0.85
40+	3	0.15	1.00

d) $s_x = 25.30259$

e) $30 - 9.5 = 20.5$

8.25 The *Los Angeles Times* (Oct. 30, 1983) reported that a typical customer of the 7-Eleven convenience stores spends $\$3.24$. Suppose that the average amount spent by customers of 7-Eleven stores is the reported value of $\$3.24$ and that the standard deviation for the amount of sale is $\$8.88$.

a) What is the level of measurement for these data?

b) Based on the given mean and standard deviation, do you think that the distribution of the variable *amount of sale* could have been symmetric in shape? Why or why not?

c) What can be said about the proportion of all customers that spend more than $\$20$ on a purchase at 7-Eleven?

a) Ratio.

b) These data are highly skewed (asymmetric) since there is a truncation point at zero on one side and it is well within the standard deviation.

c) Using Chebychev's Inequality with k = 2 (since $\mathbf{2s = 21 \approx 20}$): $\mathbf{100(1 - \frac{1}{2^2}) = 75}\%$, so at least 75% of the customers spend between $\$0.00$ and $\$21.00$, so at most 25% spend more than $\$21.00$, and the proportion that spend more than $\$20.00$ is at most a little higher than 25%. If you figured the exact k, thats great but not necessary for credit.

9

Markov Chains

9.1 Consider a lone knight on a chessboard making moves uniformly randomly from those legally available to it at each step. Show that the path of the knight (starting in a corner) is or is not a Markov chain. If so, is it irreducible and aperiodic?

This is a Markov chain, because the probability of a move is conditional only on the current state for determining the possible draws for the next state. It is irreducible since the Knight can eventually reach any state from any other (unlike, say, a bishop). This can be proven, but it is not required here. Also, clearly it is aperiodic unless additional stipulations are made about how the probabilities are changed (i.e. no longer uniform). So as long as the uniformly random choice stays uniformly random, this is an aperiodic Markov chain.

9.3 Using this matrix:

$$\mathbf{X} = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

find the vector of stationary probabilities.

Actually, for this simple example we can solve directly for the steady state

$S = [s_1, s_2]$ by stipulating:

$$\begin{bmatrix} s_1 & s_2 \end{bmatrix} \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \end{bmatrix},$$

and solving the resulting two equations for the two unknowns. This is done by starting with:

$$[0.75s_1 + 0.5s_2, 0.25s_1 + 0.5s_2] = [s_1, s_2]$$

which produces:

$$0.5s_2 = 0.25s_1, \qquad 0.25s_1 = 0.5s_2.$$

It seems like we are stuck here since these produce the same information, but the second (needed) equation comes from the fact that there are only two states so the probabilities must sum to one: $s_1 + s_2 = 1$. Plugging $s_1 = 1 - s_2$ into $0.5s_2 = 0.25s_1$ produces $s_2 = \frac{1}{3}$. We therefore also know that $s_1 = \frac{2}{3}$.

9.5 There are many applications and famous problems for Markov chains that are related to gambling. For example, suppose a gambler bets \$1 on successive games unless she has won three in a row. In the latter case she bets \$3 but returns to \$1 if this game is lost. Does this dependency on more than the last value ruin the Markovian property? Can this process be made to depend only on a previous "event"?

We can redefine the chain in the following way, starting with the fourth bet, lump the last three iterations of the game into one "super-iteration." In this way there are now only two states within each of these new steps according to:

- STATE 1: less than three wins in the last three plays.
- STATE 2: three wins in the last three plays.

Therefore there are two betting decisions:

- STATE 1: bet \$1.
- STATE 2: bet \$3.

Since the probabilities do not change as the game is played and betting changes, there are no additional complications to keep track of. So the new chain has a "moving window" of three plays over the previous chain. Now the probability of moving to a new state is conditional only on the last state under the new definition.

9.7 Consider the prototypical example of a Markov chain kernel:

$$\begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix},$$

where $0 \leq p, q \leq 1$. What is the stationary distribution of this Markov chain? What happens when $p = q = 1$?

To get the stationary distribution start with the definition:

$$\begin{bmatrix} s_1 & s_2 \end{bmatrix} \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \end{bmatrix},$$

So now:

$$s_1(1-p) + s_2 q = s_2$$

$$(1-s_2)(1-p) + s_2 q = s_2$$

$$1 - p - s_2 + s_2 p + s_2 q = s_2$$

$$s_2 = \frac{1-p}{2-p-q}.$$

Therefore $s_1 = \frac{1-q}{2-q-p}$. If $p = q = 1$ this stationary distribution does not exist (see the denominator). Under this condition this is not a stochastic process since the starting state is permanent (absorbing).

9.9 Suppose that for a Congressional race the probability that candidate B airs negative campaign advertisements in the next period, given that candidate A has in the current period, is 0.7; otherwise it is only 0.07. The same probabilities apply in the opposite direction. Answer the following questions.

a) Provide the transition matrix.

b) If candidate B airs negative ads in period 1, what is the probability that candidate A airs negative ads in period 3?

c) What is the limiting distribution?

First of all define the two-dimensional state with a vector with positions for candidate A and candidate B, where a 0 means no negative ads and a 1 means negative ads. Therefore the state space consists of $\{(0,0),(0,1),(1,0),(1,1)\}$, where the first position corresponds to A and the second corresponds to B. Now we need to keep track of joint probabilities since two things are happening (supposedly independently) at once. So the probability of transitioning from $(0,1)$ to $(0,0)$, for instance, is the probability that A stays non-negative when B is negative, $1-0.7$, times the probability that B goes non-negative for a non-negative A in the previous period, 0.93. The probability of this particular transition is therefore 0.291. Note that an assumption of this setup is the decisions are based only on the other actor's last behavior not ones' own.

a) So for each of these scenarios we need to calculate the corresponding joint probability. This is done in the table below where the columns indicate states at time $t-1$ and the rows indicate potential states at time t.

$$\begin{array}{c} \\ (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{array} \begin{array}{c} \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \end{array} \\ \begin{pmatrix} 0.865 & 0.279 & 0.279 & 0.09 \\ 0.065 & 0.021 & 0.651 & 0.21 \\ 0.065 & 0.651 & 0.021 & 0.21 \\ 0.005 & 0.049 & 0.049 & 0.49 \end{pmatrix} \end{array},$$

Notice that the columns sum to one.

b) If candidate B airs negative ads in period 1, the probability that candidate A airs negative ads in period 3 is determined by stipulating an appropriate starting value and running the chain two iterations by multiplying by the transition kernel (the transpose of the matrix above). To satisfy the condition, we can start from either $(0, 1)$ or $(1, 1)$, so it is necessary to evaluate both.

Starting at $(0, 1)$, we get the following probabilities:

Scenario	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
Iteration 1	0.43323	0.45267	0.05577	0.05833
Iteration 2	0.43323	0.45267	0.05577	0.05833

So the probability that candidate A airs negative ads in period 3 is the sum of the probabilities of scenario 3 and 4 at Iteration 2: $0.05577 + 0.05833 = 0.11410$.

Starting at $(1, 1)$, we get the following probabilities:

Scenario	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
Iteration 1	0.23913	0.24987	0.24987	0.26113
Iteration 2	0.23913	0.24987	0.24987	0.26113

Again, the probability that candidate A airs negative ads in period 3 is the sum of the probabilities of scenario 3 and 4 at Iteration 2: $0.24987 + 0.26113 = 0.51100$.

Now if we want to say that there is a 0.5 probability of candidate A being zero or one at the start, then we simply take the mean of these to get an expected probability: $\frac{1}{2}(0.11410) + \frac{1}{2}(0.51100) = 0.312550$.

c) The limiting distribution can be found by running the chain as above, but not stopping until the probabilities are stable (we could also do this analytically but it is a lot more work). After about 40 iterations there are no changes in the observed probabilities (with the accuracy of `R`), producing:

Scenario	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
Prob.	0.657523	0.153288	0.153288	0.0359016

9.11 Given an example transition matrix that produces a *non*irreducible Markov chain, and show that it has at least two distinct limiting distributions.

These actually occur in growth models in economics and are an interesting area of study. Below is a unimaginative example where the starting point determines the final limiting distribution.

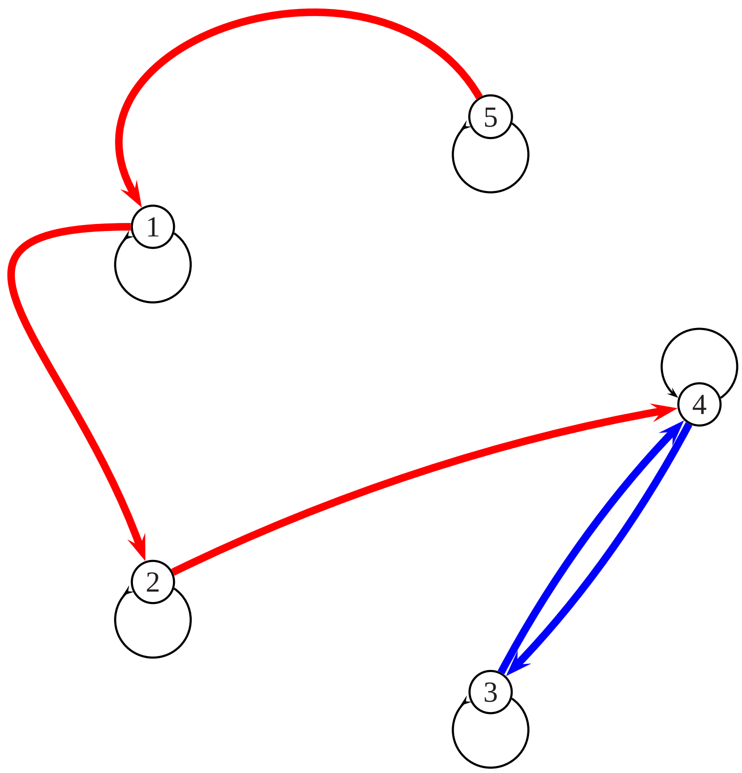
$$\begin{bmatrix} 0.2 & 0.8 & 0 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.3 & 0.6 \\ 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0.2 & 0.3 & 0.5 \end{bmatrix}$$

If the chain is started in states 1 or 2, then the limiting distribution is $[0.529, 0.0.471, 0, 0, 0]$, and if the chain is started in states 3, 4, or 5, the limiting distribution is $[0, 0, 0.109, 0.760, 0.13]$.

9.13 For the following transition matrix, which classes are closed?

$$\begin{bmatrix} 0.50 & 0.50 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.75 & 0.25 & 0.00 \\ 0.50 & 0.00 & 0.00 & 0.00 & 0.50 \end{bmatrix}.$$

First, let's look at diagram of the potential paths of the chain:



A closed class is one where the Markov chain cannot leave once it enters. The closed class here is that defined by $\{3, 4\}$ since the chain will just cycle back and forth between these two states. Note that the chain is guaranteed to enter this closed class for any starting point. While there is always a positive probability of staying at any of $\{1, 2, 5\}$, the chain will eventually leave and start the cycle towards the closed class.

9.15 Consider the following Markov chain from Atchadé and Rosenthal (2005). For the discrete state space $\Theta = \{1, 3, 4\}$, at the nth step produce the $n+1^{st}$ value by:

- if the last move was a rejection, generate $\theta' \sim$ uniform$(\theta_n - 1 : \theta_n + 1)$;
- if the last move was an acceptance, generate $\theta' \sim$ uniform$(\theta_n - 2 : \theta_n + 2)$;
- if $\theta' \in \Theta$, accept θ' as θ_{n+1}, otherwise reject and set θ_n as θ_{n+1};

where these uniform distributions are on the inclusive positive integers and some arbitrary starting point θ_0 (with no previous acceptance) is assumed. What happens to this chain in the long run?

Consider the following R code for implementing this Markov chain.

```
x <- c(3); accept <- 0; num.sims <- 1000000
for (i in 1:num.sims)  {
  if (accept == 1)
    theta <- sample(size=1,(x[i]-2):(x[i]+2))
  else theta<-sample(size=1,(x[i]-1):(x[i]+1))
  if (theta == 1 || theta == 3 || theta == 4) {
    x <- c(x,theta)
    accept <- 1
  }
  else {
    x <- c(x,x[i])
    accept <- 0
  }
}
```

Running this chain for a million iterations, throwing away the first $100,000$ produces a the distribution described by the following table.

1	3	4
423376	243911	232713

Interestingly, the mass is nearly twice as much on 1.

9.17 Dobson and Meeter (1974) modeled the movement of party identification in the United States between the two major parties as a Markovian. The following transition matrix gives the probabilities of *not* moving from one status to another conditional on moving (hence the zeros on the diagonal):

$$\begin{array}{c} \\ SD \\ WD \\ I \\ WR \\ SR \end{array} \begin{array}{c} \begin{array}{ccccc} SD & WD & I & WR & SR \end{array} \\ \left(\begin{array}{ccccc} 0.000 & 0.873 & 0.054 & 0.000 & 0.073 \\ 0.750 & 0.000 & 0.190 & 0.060 & 0.000 \\ 0.154 & 0.434 & 0.000 & 0.287 & 0.125 \\ 0.039 & 0.205 & 0.346 & 0.000 & 0.410 \\ 0.076 & 0.083 & 0.177 & 0.709 & 0.000 \end{array}\right) \end{array},$$

where the labels indicate $SD =$ Strong Democrat, $WD =$ Weak Democrat, $I =$ Independent, $WR =$ Weak Republican, and $SR =$ Strong Republican. Convert this to a transition matrix for moving from one state to another and show that it defines an ergodic Markov chain. What is the stationary distribution?

This problem could be explained better than it is here. Dobson and Meeter say:

"The entries in the table are to be interpreted as follows. Consider the first row (i.e. 1956 Strong Democrats). The probability that an individual moved from Strong Democrat in 1956 to Weak Democrat in 1958, *given that he moved*, is .87."

So given this condition, and not knowing the relative proportion that remained at their current ideology, we can just treat this as a transition matrix for movers only if we want. The authors do provide row totals ($55, 116, 136, 78, 79$) for combined movers and stays, but this does not let us recover the proportion not moving. Furthermore, these numbers are noticeably different than what appears in the ANES file (7252 American Panel Study: 1956, 1958, 1960). So this problem can be used as an introductory data research problem on downloading or as just another exercise in calculating limiting distributions. The interpretation of the limiting distribution is a little odd since cases enter and leave the analysis by changing their status between stayers and movers. The code is below. After $1,000$ iterations the chain appears to have settled down to:

$$[42.03268, 49.92121, 20.09307, 17.77727, 12.81688]$$

```
P<-matrix(c(0.000, 0.873, 0.054, 0.000, 0.073,
  0.750, 0.000, 0.190, 0.060, 0.000,
  0.154, 0.434, 0.000, 0.287, 0.125,
  0.039, 0.205, 0.346, 0.000, 0.410,
  0.076, 0.083, 0.177, 0.709, 0.000),
  byrow=TRUE,ncol=5)

MC.multiply <- function(P.in,N)  {
    P1 <- rep(0.5,5)%*%P.in
    for (i in 1:(N-1))  {
        P1 <- P1%*%P.in
        print(P1)
    }
    P1
}

MC.multiply(P,1000)
```

We can also download and condition the ANES file to perform the same analysis on the switching numbers there but with the respondents that do not switch preserved on the diagonal:

```
library(foreign)
id.panel <- read.dta(07252-0001-Data.dta)

> levels(id.panel$V560088)
  [1] "STRONG DEMOCRAT"
  [2] "NOT VERY STRONG DEMOCRAT"
  [3] "INDEPENDENT CLOSER TO DEMOCRATS (’YES,"
```

```
  [4] "INDEPENDENT (’NO, NEITHER’ OR ’DK’ TO"
  [5] "INDEPENDENT CLOSER TO REPS (’YES, REPU"
  [6] "NOT VERY STRONG REPUBLICAN"
  [7] "STRONG REPUBLICAN"
  [8] "OTHER, MINOR PARTY AND REFUSED TO SAY"
  [9] "APOLITICAL (DK TO Q"
[10] "NA"

id.data
  <- cbind(id.panel$V560088,id.panel$V580360)
id.data2 <- NULL
for (i in 1:nrow(id.data))  {
  if ((id.data[i,1] < 8) & (id.data[i,2] < 8))
    id.data2 <- rbind(id.data2,id.data[i,])
}
for (i in 1:nrow(id.data2))  {
  if ((id.data2[i,1] == 3)
     || (id.data2[i,1] == 5))
    id.data2[i,1] <- 4
  if ((id.data2[i,2] == 3)
     || (id.data2[i,2] == 5))
    id.data2[i,2] <- 4
}

P <- table(id.data2[,1],id.data2[,2])

       1   2   4   6   7
   1 179  45   3   1   4
   2  84 134  22   5   0
   4  22  50 142  34  16
```

```
   6   3  13  24  93  27
   7   5   4  11  49 100

row.totals <- apply(P,1,sum)
Q <- sweep(P,1,row.totals,FUN="/")
```

```
            1           2           4           6           7
1 0.771551724 0.193965517 0.012931034 0.004310345 0.017241379
2 0.342857143 0.546938776 0.089795918 0.020408163 0.000000000
4 0.083333333 0.189393939 0.537878788 0.128787879 0.060606061
6 0.018750000 0.081250000 0.150000000 0.581250000 0.168750000
7 0.029585799 0.023668639 0.065088757 0.289940828 0.591715976
```

```
MC.multiply(Q,1000)
```

This settles down well before the $1,000$ iterations, giving:

$$[1.121576, 0.6503875, 0.2703534, 0.2619291, 0.1957537]$$