A3SR Math Review

Fall 2019

A Note:

The notes in this packet are not meant to be comprehensive in any way, nor should they be used for learning this material for the first time. Instead, the resources provided here are meant to accomplish two main goals: 1) to give you a sense of the skills/topics that professors will expect you to have learned prior to the start of this program and 2) to jog your memory/refresh your working knowledge of these concepts if you need a quick re-cap. For anything that you haven't learned before or have not seen in many years, the notes provided here may not be sufficient. Suggested outside resources are provided in a separate document, including links to additional practice problems. In designing these materials, we have tried to be concise, and have only included topics that are directly related to concepts covered in your first semester courses. We have also tried to find a balance between things you are expected to understand versus things that you will be expected to calculate. In this program, a lot of complex calculations can be done in R (or using other online tools, like Wolfram Alpha). Most of the time, you will not be asked to do this work by hand (at least not more than once). However, you will find it much easier to understand results in R (and potential errors) if you have a sense of what is going on "under the hood". More in-depth familiarity and experience with these topics is always helpful (if you have additional time for review), and we hope that this packet provides guidance on where to focus your attention.

Properties of Logarithms

Relevant Courses:

-Quantitative Methods
-Generalized Linear Models

Notes

Definition

Logarithms are defined such that $log_b(A) = X$ is equivalent to $b^X = A$

Properties

Using properties of exponents and the definition above, we can derive the following:

1. The Product Rule: $log_b(MN) = log_b(M) + log_b(N)$

2. The Quotient Rule: $log_b(\frac{M}{N}) = log_b(M) - log_b(N)$

3. The Power Rule: $log_b(M^p) = plog_b(M)$

4. $log_b(b^X) = X$

5. $b^{log_b(X)} = X$

6. $log_b b = 1$

7. $log_b 1 = 0$

Example 1: Expanding logarithms

$log_e(\frac{2x^3}{y})$
$= log_e(2x^3) - log_e(y)$
$= log_e(2) + log_e(x^3) - log_e(y)$
$= log_e(2) + 3log_e(x) - log_e(y)$

Example 2: Condensing logarithms

$2log_3(x) + log_3(5) - log_3(2)$
$= log_3(x^2) + log_3(5) - log_3(2)$
$= log_3(5x^2) - log_3(2)$
$= log_3(\frac{5x^2}{2})$

Practice Problems

1. Solve the following:
   a. $log_e(e^x)$
   b. $log_{10}(100)$
   c. $log_{10}(\frac{1}{10})$
   d. $log_{10}(0)$

2. Expand the following:
   a. $log_{10}(\frac{5y^3}{x})$
   b. $log_2(\frac{4y^2}{3x})$
   c. $log_e(2x^2y^3)$

3. Condense the following:
   a. $4log_3(x) - 2log_3(y)$
   b. $log_2(x) + 5log_2(y) - log_2(5)$
   c. $log_{10}(5) + log_{10}(2)$

Answers

1. Solve:
   a. x
   b. 2
   c. -1
   d. There is no solution because there is no power of 10 that would equal 0

2. Expand:
   a. $log_{10}(5) + 3log_{10}(y) - log_{10}(x)$
   b. $2 + 2log_2(y) - log_2(3) - log_2(x)$
   c. $log_e(2) + 2log_e(x) + 3log_e(y)$

3. Condense
   a. $log_3(\frac{x^4}{y^2})$
   b. $log_2(\frac{xy^5}{5})$
   c. 1

Matrix Algebra

Relevant Courses:

-Quantitative Methods

Notes

Definitions

An $m \times n$ matrix A has m rows, n columns, and can be written as:

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix}$$

The transpose of an $m \times n$ matrix, A is the $n \times m$ matrix (denoted A^T) such that every element a_{ij} in matrix A is moved to row j and column i. For example, if:

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

then,

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 5 & 7 \end{bmatrix}$$

The $n \times n$ identity matrix I_n is a matrix with 1s on the diagonal and 0s everywhere else:

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

An $n \times n$ matrix is called "diagonal" if all elements not on the diagonal are zeros. For example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

An $n \times n$ matrix is called "upper triangular" if all elements below the diagonal are zeros. For example:

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

An $n \times n$ matrix is called "lower triangular" if all elements above the diagonal are zeros. For example:

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

Suppose that an we have 2 $n \times n$ matrices, A and B, such that $AB = I_n$ (note: this also implies $BA = I_n$). Then we say that B is the inverse of A (and vice versa) and we can write, $B = A^{-1}$. A matrix A has an inverse if and only if its determinant is not equal to zero. Note that the determinant of a 2×2 matrix can be calculated as follows (it is not important that you are able to calculate the determinant of a higher dimensional matrix by hand):

$$det(\begin{bmatrix} a & b \\ c & d \end{bmatrix}) = ad - bc$$

Operations with matrices

a. Adding matrices ($A + B = C$): If we add 2 $m \times n$ matrices, A and B, we get another $m \times n$ matrix C such that $c_{ij} = a_{ij} + b_{ij}$

b. Subtracting matrices ($A - B = C$): If we subtract the $m \times n$ matrix B from the $m \times n$ matrix A, we get another $m \times n$ matrix C such that $c_{ij} = a_{ij} - b_{ij}$

c. Multiplying a matrix by a scalar ($cA = B$): If we multiply the $m \times n$ matrix A by a scalar, c, then we get another $m \times n$ matrix B such that $b_{ij} = c * a_{ij}$

d. Matrix multiplication ($AB = C$): Note that it is only possible to compute AB if the number of columns in matrix A equals the number of rows in matrix B. If this is the case, then when we multiply an $m \times n$ matrix A by an $n \times k$ matrix B, we get an $m \times k$ matrix, C, such that $c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}$

Example 1: Matrix Multiplication

$$\begin{bmatrix} 3 & 4 \\ 2 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 7 & 2 & 1 \\ 3 & 5 & 2 \end{bmatrix} = \begin{bmatrix} (3*7+4*3) & (3*2+4*5) & (3*1+4*2) \\ (2*7+5*3) & (2*2+5*5) & (2*1+5*2) \\ (1*7+2*3) & (1*2+2*5) & (1*1+2*2) \end{bmatrix} = \begin{bmatrix} 33 & 26 & 11 \\ 29 & 29 & 12 \\ 13 & 12 & 5 \end{bmatrix}$$

Properties of matrix operations

1. $A + B = B + A$

2. $(A + B) + C = A + (B + C)$

3. $(AB)C = A(BC)$

4. $(A + B)C = AC + BC$

5. If A is an $m \times n$ matrix, then $I_m A = A$ and $AI_n = A$

Note that, in general $AB \neq BA$

Writing a system of equations using matrix notation

Note that, if we have a system of equations:
$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$
$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n$
$\cdots$
$y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n$

We can re-write these equations much more simply as:
$Y = AX$ where Y is a $1 \times m$ matrix, A is an $m \times n$ matrix, and X is a $1 \times n$ matrix:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

Practice Problems

1. Solve the following:

   a. $\begin{bmatrix} 2 & 4 & 2 \\ 1 & 3 & 0 \\ 1 & 6 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 5 & 0 \\ -2 & -3 & 0 \\ 1 & 9 & 5 \end{bmatrix}$

   b. $\begin{bmatrix} 2 & 1 \\ -2 & 2 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 5 & 2 & -1 \\ 3 & 4 & 2 \end{bmatrix}$

   c. Let A$=\begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 4 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 4 \\ 1 & 2 \\ 7 & 0 \end{bmatrix}$ Calculate $A^T B$

   d. Using the same matrices as in part c, calculate $B^T A$

2. Show that A and B are inverses:

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

   $B = \begin{bmatrix} -1 & 0.5 & 1 \\ 1 & 0 & -1 \\ 0 & -0.5 & 1 \end{bmatrix}$

3. Suppose that A is a 4x3 matrix and B is a $3x8$ matrix.
   a. Does AB exist? If so, what are the dimensions of AB?
   b. Does BA exist? If so, what are the dimensions of BA?

4. What is the determinant of $\begin{bmatrix} 1 & -2 \\ 4 & 3 \end{bmatrix}$?

Answers

1. Solve:

   a. $\begin{bmatrix} 3 & 9 & 2 \\ -1 & 0 & 0 \\ 2 & 15 & 7 \end{bmatrix}$

   b. $\begin{bmatrix} 13 & 8 & 0 \\ -4 & 4 & 6 \\ 26 & 16 & 0 \end{bmatrix}$

   c. $\begin{bmatrix} 35 & 10 \\ 13 & 18 \end{bmatrix}$

   d. $\begin{bmatrix} 35 & 13 \\ 10 & 18 \end{bmatrix}$

2. Show that A and B are inverses:

   $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0.5 & 1 \\ 1 & 0 & -1 \\ 0 & -0.5 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

3. Suppose that A is a 4x3 matrix and B is a $3x8$ matrix.
   a. AB exists and is 4x8
   b. BA does not exist

4. The determinant is $(1 * 3) - (-2 * 4) = (3) - (-8) = 11$

Derivatives

Relevant Courses:

-Probability
-Quantitative Methods

Notes

Definition

The derivative of a function $y = f(x)$ with respect to x is defined as a function giving the instantaneous slope of $y = f(x)$ for any value x. Notationally, a derivative can be written in any of the following ways:

$$f'(x) = y' = \frac{df}{dx} = \frac{d}{dx}(f(x)) = \frac{dy}{dx} = \frac{d}{dx}(y)$$

The second derivative of $y = f(x)$ with respect to x is a function giving the rate of change of the instantaneous slope of $f(x)$. Notationally, it can be represented in any of the following ways (note: 3rd, 4th, etc. derivatives are notated in a similar way, with increasing exponents or $'$s):

$$f''(x) = y'' = \frac{d^2f}{dx^2} = \frac{d^2}{dx^2}(f(x)) = \frac{d^2y}{dx^2} = \frac{d^2}{dx^2}(y)$$

Using derivatives to find local minima and maxima of a function

To find all local minima or maxima of a function $y = f(x)$:

1. Take the first derivative of $f(x)$ with respect to x ($f'(x)$).
2. Set this expression equal to zero and solve for x. These values of x are local maxima and minima.
3. Calculate the second derivative of $f(x)$ with respect to x ($f''(x)$)
4. Plug in the values of x calculated in part 2. If the second derivative is positive, this value of x represents a local minimum; if the second derivative is negative, it is a local maximum

Properties

Properties of derivatives:

1. Sum/Difference rule: $(f(x) \pm g(x))' = f'(x) \pm g'(x)$
2. Constant multiple rule: $(cf(x))' = cf'(x)$ where c is a constant
3. Power rule: If $f(x) = x^n$, then $f'(x) = nx^{n-1}$
4. Product rule: $(f(x)g(x))' = f'(x)g(x) + g'(x)f(x)$
5. Quotient rule (given g(x) != 0): $(\frac{f(x)}{g(x)})' = \frac{f'(x)g(x)-g'(x)f(x)}{g^2(x)}$
6. Chain rule: $(f(g(x)))' = f'(g(x)) * g'(x)$

Partial derivatives

For a function of more than one variable (i.e., $f(x,y)$), we can take partial derivatives with respect to each variable. The partial derivative of $f(x,y)$ with respect to x is often denoted by either $\frac{\partial f}{\partial x} = f_x$. The partial derivative with respect to y would be denoted by $\frac{\partial f}{\partial y} = f_y$. The partial derivative with respect to x would be calculated by treating any non-x variables as constants when applying the above properties.

Practice Problems

1. Find $f'(x)$ for each of the following. Then compute $f''(x)$ for a-d:
   a. $f(x) = 5x^2 + 3x + 1$
   b. $f(x) = \frac{5}{x^3} + 2x^4$
   c. $f(x) = (3x+1)^5$
   d. $f(x) = 2x^3(x^2+1)$
   e. $f(x) = \frac{2x+1}{x^2-5}$
2. Find the partial derivative with respect to x for each of the following:
   a. $f(x,y) = 3xy^2 + 2x$
   b. $f(x,y) = (xy^4 + 2y)^3$
   c. $f(x,y) = 4x^3 + xy + x^2y^2 + 4x + 2$
3. Find all local minima and maxima for the following function (and note whether they are minima or maxima): $\frac{2}{3}x^3 - x^2 - 12x$

Solutions

1. $f'(x)$ and $f''(x)$
   a. $f'(x) = 10x + 3$ and $f''(x) = 10$
   b. $f'(x) = \frac{-15}{x^4} + 8x^3$ and $f''(x) = \frac{-60}{x^5} + 24x^2$
   c. $f'(x) = 15(3x+1)^4$ and $f''(x) = 180(3x+1)^3$
   d. $f'(x) = 10x^4 + 6x^2$ and $f''(x) = 40x^3 + 12x$
   e. $f'(x) = \frac{2(x^2-5)-2x(2x+1)}{(x^2-5)^2}$ and $f''(x)$
2. $f_x(x,y) =$
   a. $3y^2 + 2$
   b. $3y^4(xy^4 + 2y)^2$
   c. $12x^2 + y + 2xy^2 + 4$
3. Local minima and maxima
   a. $f'(x) = 2x^2 - 2x - 12 = (2x+4)(x-3)$, so setting the first derivative equal to 0, we get $0 = (2x+4)(x-3)$, with solutions $x = 3$ and $x = -2$. $f''(x) = 4x - 2$. $f''(3) = 10$ and $f''(x) = -10$. So, $x = 3$ is a local minimum and $x = -2$ is a local maximum. $f(3) = 18 - 9 - 36 = -27$ and $f(-2) = \frac{-16}{3} - 4 + 24 = \frac{44}{3} \approx 14.67$

Integrals

Relevant Courses:

-Probability
-Quantitative Methods

Notes

Definition

Indefinite integrals

In the previous section, we calculated the derivative of a function, $f(x)$. Finding an indefinite integral (also called an anti-derivative) involves a simple reversal of this process. The indefinite integral, $F(x)$, of a function, $f(x)$, is defined such that $F'(x) = f(x)$ and is written as follows:

$$\int f(x)dx = F(x) + C$$

where C is a constant.

Definite integrals

The definite integral of $f(x)$ from a to b gives the area under the curve of $f(x)$ on the interval between a and b and is calculated/notated as follows:

$$\int_a^b f(x)dx = F(b) - F(a)$$

Basic properties

Note: There are many other properties of integrals, which may be useful once or twice throughout this program. The following are what you will see most commonly, but you should feel comfortable looking up and using additional properties as needed.

1. $\int cf(x)dx = c\int f(x)dx$ where c is a constant
2. $\int f(x) \pm g(x)dx = \int f(x)dx \pm \int g(x)dx$
3. $\int \frac{1}{x}dx = ln|x| + C$
4. $\int x^n dx = \frac{1}{n+1}x^{n+1} + C$

U substitution

Practice Problems

1. Evaluate the following indefinite integrals:
   a. $\int 5x^2 dx$
   b. $\int 2x^3 - \frac{3}{x^2} dx$
   c. $\int (2x-2)(x+3) dx$
   d. $\int \frac{2}{x} dx$
2. Evaluate the following definite integrals:
   a. $\int_0^1 4x^3 dx$
   b. $\int_{-2}^2 8x^3 + 3x^2 - 5x + 2dx$
   c. $\int_1^\infty \frac{2}{x^2} dx$
   d. $\int_1^2 \int_0^2 2x^3 + 3y^3x^2 \ dxdy$

Answers

1. Indefinite integrals:
   a. $\int 5x^2 dx = \frac{5}{3}x^3$
   b. $\int 2x^3 - \frac{3}{x^2} dx = \int 2x^3 - 3x^{-2} dx = \frac{1}{2}x^4 + 3x^{-1} + C$
   c. $\int (2x-2)(x+3) dx = \int 2x^2 + 4x - 6dx = \frac{2}{3}x^3 + 2x^2 - 6x + C$
   d. $\int \frac{2}{x} dx = 2 \int \frac{1}{x} dx = 2ln|x| + C$
2. Definite integrals:
   a. $\int_0^1 4x^3 dx = x^4|_0^1 = 1^4 - 0^4 = 1$
   b. $\int_{-2}^2 8x^3 + 3x^2 - 5x + 2dx = (2x^4 + x^3 - \frac{5}{2}x^2 + 2x)|_{-2}^2 = (32 + 8 - 10 + 4) - (32 - 8 - 10 - 4) = 24$
   c. $\int_1^\infty \frac{2}{x^2} dx = \int_1^\infty 2x^{-2} dx = -2x^{-1}|_1^\infty = 0 - -2 = 2$
   d. $\int_1^2 \int_0^2 2x^3 + 3y^3x^2 \ dxdy = \int_1^2 (\frac{1}{2}x^4 + y^3x^3)|_0^2 dy = \int_1^2 8 + 8y^3 dy = (8y + 2y^4)|_1^2 = (16+32) - (8+2) = 38$

Variables: Types and Summaries

Relevant Courses:

-Probability
-Quantitative Methods
-Statistical Computing

Notes

Types of variables:

1. Numerical
   a. continuous (example: height)
   b. discrete (example: population)

2. Categorical
   a. nominal (unordered) (example: responses to: "who are you planning to vote for in the upcoming election?")
   b. ordinal (ordered) (example: responses to: "On a scale from 1-10, how happy are you right now?")

Summarizing variables:

Note: without going into too much detail, it is important that summary statistics be chosen and evaluated with the variable type in mind. For example, if I calculate the mean of a variable which is categorical (nominal) with responses coded as 0s and 1s, the mean will give me the percentage of 1s in my dataset. However, this is no longer the case if the same categorical variable is coded with 1s and 2s.

Summary statistics for central tendency

1. Mean: For a variable x and sample size n, the mean is given by $\frac{\Sigma_{i=1}^{n} x_i}{n}$

2. Median: For a variable x, the median is defined as the central value at which 50% of x-values are smaller and 50% are larger

3. Mode: For a variable x, the mode is the most common value of x. A variable can have multiple modes

Summary statistics for spread

1. Range: For a variable x, the range is given by max(x)-min(x)

2. Population Variance: For a variable x and population size n, the population variance is given by $\sigma^2 = \frac{\Sigma_{i=1}^{n}(x_i - \mu)^2}{n}$ where μ is the mean (i.e., average squared difference between each value of x and the mean).

3. When using a sample to estimate the variance of x in a larger population, sample variance is calculated (ever so slightly differently) as: $\sigma^2 = \frac{\Sigma_{i=1}^{n}(x_i - \mu)^2}{n-1}$.

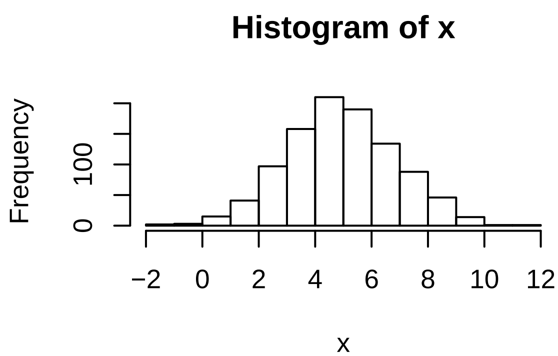
4. Standard deviation: $\sigma = \sqrt{\sigma^2}$

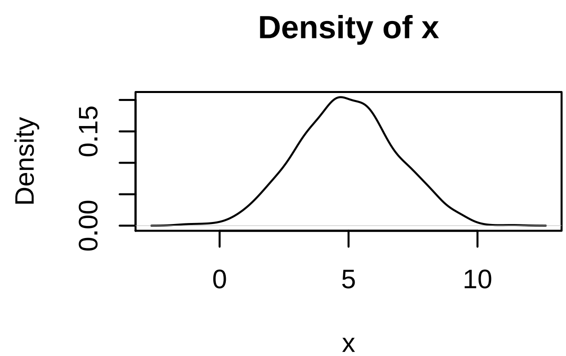
A note about population variance vs. sample variance: In many applications of statistics, we examine a sample of data from a larger population of interest, and try to make inference about the larger population. For example, if we wanted to estimate the mean height of adult women, we might find a sample of 1000 adult women, calculate the mean and variance of their heights and say that these estimates (approximately) also apply to the larger population (we can be more confident about this for larger sample sizes). However, (and this may sound strange), suppose we knew the heights of all adult women and could calculate the variance in their heights. Then suppose we took a bunch of random samples of 1000 women from this population and calculated the variance in heights for each of these samples, using the same formula as we did for population variance. It turns out that these sample variances would, on average, be slightly off from the true population variance, even as the number of samples goes to infinity. Thus, the formula for population variance is actually a "biased" estimator of population variance when applied to a sample. Therefore, we use a slightly different formula for sample variance to "fix" this issue.

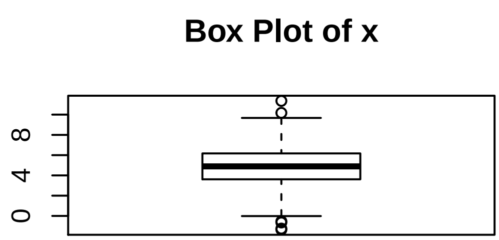
Another note: You might see the term "standard error" thrown around and wonder how this is different from standard deviation. This may be worth a quick Google search if it is unfamiliar (and will be briefly addressed in a later section introducing the central limit theorem).

Summarizing variables visually

A few common ways of visualizing data (including central tendency, spread, and other characteristics like bimodality and skewness) are demonstrated below (histograms, smoothed density, and box plots). These are worth looking up if they are not immediately familiar.







Basic Probability

Relevant Courses:

-Probability

Notes

Random process: Something that produces observation(s) with some level of uncertainty.
Sample point: A single possible outcome of a random process.
Sample space: All possible sample points for an random process.

Here are some examples of random processes, sample points, and sample spaces:

a) Random process: Roll a standard 6-sided die and record the result (the number that is face up)
example of a sample point: 2
sample space: {1,2,3,4,5,6}

b) Random process: Roll a die, then flip a coin (with probability .6 of landing on heads) and record the die roll and whether heads or tails is facing upward on the coin.
example of a sample point: (2,Heads)
sample space: {(1,Heads),(2,Heads),(3,Heads),(4,Heads),(5,Heads),(6,Heads),(1,Tails),(2,Tails),(3,Tails),(4,Tails),(5,Tails),(6,Tails)}

Probability Definition

Suppose that a random process is repeated infinitely many times; Then the probability of a particular outcome is the proportion of times that the outcome is observed.

Law of Large Numbers

As a random process is repeated more times, the observed proportion of any particular outcome will converge to the true empirical probability of that outcome

Addition Rule

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
Note that if A and B are disjoint then $P(A \text{ and } B) = 0$

Complements

For a particular outcome, A, of a random process, the complement of that outcome, A^c is made up of all sample points *not* in A.
$A + A^c = 1$

Multiplication Rule (for independent events)

If A and B are independent:
$P(A \text{ and } B) = P(A) \times P(B)$

Random Variables and Probability Density/Mass Functions

Relevant Courses:

-Quantitative Methods
-Probability
-Statistical Computing -Causal Inference

Notes

Random variables are functions with *numerical* outcomes that occur with some level of uncertainty. Often, experiments with non-numerical outcomes are represented by random variables. For example, flipping a fair coin one time could be considered a random variable with two possible outcomes: 0 (corresponding to tails) or 1 (corresponding to heads).

Random variables can be discrete or continuous:

1. Discrete random variables have a finite number of potential outcomes

2. Continuous random variables have infinitely (countably or uncountably) many possible outcomes

A probability distribution of a random variable determines the relative likelihood of each potential outcome of that random variable.

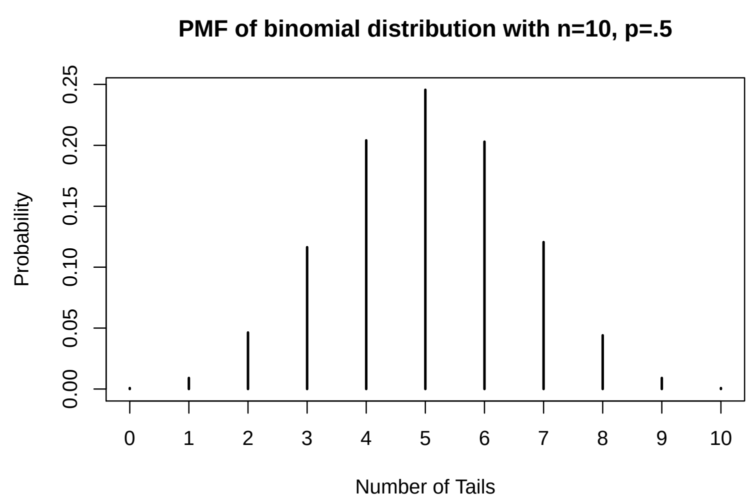
1. Probability Mass Function (PMF): (for discrete random variables) maps possible values of a random variable to their probabilities.

2. Probability Density Function (PDF): (for continuous random variables) area under the curve between any two values gives the probability of the random variable's outcome being between those two values.

There are many common probability distributions that you will need to know (these are mostly covered in Probability), but two particularly common ones are given as examples below. (Note: the major distributions covered in Probability are Bernoulli, Binomial, Normal, Uniform, Geometric, Negative Binomial, Hypergeometric, Poisson, Exponential, Gamma, Beta, Chi-Square, and Student-t. These are important to review if you are not taking Probability, and are worth looking up if you have never seen them before).

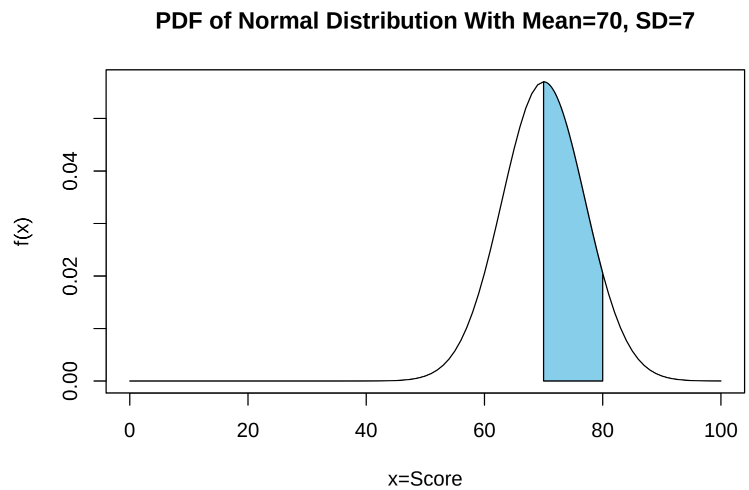
PMF Example

Binomial(n=10,p=.5) distribution: Can be thought of as the probability of x tails among 10 flips of a fair coin



PDF Example

Another common probability distribution is the normal distribution. Suppose that you are told that scores on a particular test are (approximately) normally distributed with mean 70 and standard deviation 7 (for the purposes of this thought experiment, suppose that test scores are continuous, not discrete). Then, the area of the blue shaded area in the distribution below would give the approximate probability of any particular student earning between a 70% and 80%.



Note:
It is helpful to remember the following for any data that follows a normal distribution:
-Approximately 68% of the data falls within 1 standard deviation of the mean
-Approximately 95.5% of the data falls within 2 standard deviations of the mean
-More than 99% of the data falls within 3 standard deviations of the mean

Z-Scores

Often, we want to compare two numbers and understand the extent to which they are different. One way to do this is by using Z-scores. Z-scores are generally used for observations that follow an approximately normal distribution. The Z-score for a particular observation x from a distribution with mean μ and standard deviation σ is given by $Z = \frac{x-\mu}{\sigma}$. The Z-score for any observation tells you how far (in standard deviations) that observation is from the mean, μ. If the observation came from a normal distribution, then Z-scores follow a standard ($\mu = 1, \sigma = 1$) normal distribution and a Z-table can be used to convert the Z-Score to a percentile (or vice versa). You will learn to use a Z-table in probability, but you might want to look this up if you have never seen it before.

Example

Suppose test scores on a particular exam are normally distributed with mean 70 and standard deviation 10. Mike gets a score of 60 and Sarah gets a score of 90. Then Mike's Z-score would be $\frac{60-70}{10} = -1$, indicating that he scored 1 standard deviation below average. Using a Z-table, we could determine that this is about 16th percentile. Meanwhile, Sarah's Z-score is $\frac{90-70}{10} = 2$, indicating that she scored 2 standard deviations above average, which is approximately the 98th percentile

Central Limit Theorem Introduction

Relevant Courses:

-Quantitative Methods
-Statistical Computing
-Probability -Causal Inference

Notes

Much of statistics is based on taking a sample from a population and using that sample to make inference about the larger population.

Suppose that we want to know the average salary of "Data Scientists". To estimate this average salary, we find a sample of 1000 people whose title is "Data Scientist" and we ask them to report their salary. Then, we take the mean of the resulting salaries and we want to know how confident we are that this mean is representative of the larger population. To do that, we can leverage the central limit theorem, which says (essentially; I am leaving out a few caveats):

If we take repeated samples of size n from a population of interest and use the sample mean (denoted $\bar{x}$) to estimate the population mean (denoted μ), then these sample means will be (approximately, if sample size is small) normally distributed with mean μ and standard deviation (which we actually call the *standard error* of $\bar{x}$) $\frac{\sigma}{\sqrt{n}}$ where σ is the population standard deviation (which we can estimate using a sample standard deviation if we don't know it).

Z-Tests, T-Tests, and P-Values

Relevant Courses:

-Quantitative Methods
-Statistical Computing
-Causal Inference

Notes

Confidence Intervals

Z-Tests, T-Tests and P-Values

Correlation and Covariance

Relevant Courses:

-Quantitative Methods
-Probability

Notes

Simple Ordinary Least Squares Regression

Relevant Courses:

-Quantitative Methods