Useful Formulas

For data $\{x_1, x_2, ..., x_n\},\$

$$s^{2} = \frac{\sum_{i=1}^{n} (x_{i}^{2}) - n(\bar{x})^{2}}{n-1} = \frac{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2}}{n-1}$$

For a binomial random variable w,

$$\mathbb{E}(w) = \mu = np, \qquad \sigma = \sqrt{np(1-p)}, \qquad \mathbb{P}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

For a finite random variable x with possible values $\{x_1, ..., x_n\}$,

$$\mathbb{E}(x) = \sum_{i=1}^{n} x_i \mathbb{P}(x_i)$$

Problem 1 - (15 total points) Suppose the random variable w represents the quantity of money won in a bet, and follows the distribution

$$\mathbb{P}(w = \$10) = \frac{1}{4}, \quad \mathbb{P}(w = \$20) = \frac{1}{2}, \quad \mathbb{P}(w = -\$30) = \frac{1}{4}$$

where the "-\$30" represents a loss of \$30.

(a) (3 points) Use the definition of expected value to find $\mathbb{E}(w)$. Is this bet worth taking? Explain.

Solution: $\mathbb{E}(w) = 10\frac{1}{4} + 20\frac{1}{2} - 30\frac{1}{4} = 5$. On average you'd win \$5 !

(b) (2 points) Define a success as winning \$10 or \$20. What is the probability of a success?

Solution: $\mathbb{P}(10 \text{ or } 20) = \mathbb{P}(10) + \mathbb{P}(20) = \frac{3}{4}.$

- (c) Suppose you played this game 10 times and let the random variable x be the number of successes.
 - (i) (5 points) What is the probability of winning 9 or fewer games ? (Hint: use the binomial distribution.)

Solution: x is binomial with $n = 10, p = \frac{3}{4}$. Thus,

$$\mathbb{P}(x \le 9) = 1 - \mathbb{P}(x = 10) = 1 - \binom{10}{10} \left(\frac{3}{4}\right)^{10} \left(\frac{1}{4}\right)^0 = 1 - \left(\frac{3}{4}\right)^{10}$$

(ii) (5 points) What is the expected number of wins (expected value of x)?

Solution: $\mathbb{E}(x) = np = 10 \cdot \frac{3}{4} = \frac{15}{2} = 7.5$

Problem 2 - (14 total points) Consider the following list of data: $\{-2, -2, -1, -1, -1, 0, 1, 2, 2, 2\}$. Compute the following values from the data.

(a) (2 points) The mean (show your work).

Solution: mean = $\frac{-2-2-1-1-1+0+1+2+2+2}{10} = \frac{0}{10} = 0.$

(b) (4 points) The median (explain your reasoning clearly).

Solution: The median is the middle value of the ordered data, or the number between the middle 2 values for an even number of points. Here the median is between 0 and -1, so the median = -0.5.

(c) (2 points) The mode (explain your reasoning clearly).

Solution: The mode is the most frequent number. Here there are two modes: -1 and 2, as they each occur three times.

(d) (4 points) The variance (show your work).

Solution: Here, we first of the equations given above. To do this we first compute the mean (\bar{x}) , then the sum of the squared data values.

$$\sum_{i=1}^{n} (x_i^2) = 4 + 4 + 1 + 1 + 1 + 0 + 1 + 4 + 4 + 4 = 24$$

This gives

$$s^{2} = \frac{\sum_{i=1}^{n} (x_{i}^{2}) - n(\bar{x})^{2}}{n-1} = \frac{24 - 10(0)^{2}}{10 - 1} = \frac{24}{9} = \frac{8}{3} = 2.666...$$

(e) (2 points) The standard deviation (you guessed it, show your work!).

Solution: The standard deviation is the square root of the variance, thus $s = \sqrt{\frac{8}{3}}$.

Problem 3 - (16 total points) Let x be a normal random variable with mean 1 and standard deviation 2. Consider the following set:

$$A = \{-1 \le x \le 1\}$$
$$B = \{ 0 < x < 2\}.$$

(a) (5 points) Find $\mathbb{P}(A)$

Solution:
$$\mathbb{P}(A) = \mathbb{P}\left(\frac{-1-1}{2} \le \frac{x-1}{2} \le \frac{1-1}{2}\right) = \mathbb{P}(-1 \le z \le 0) = \frac{1}{2} - .1587 = .3413$$

(b) (6 points) Find $\mathbb{P}(A \cup B)$

Solution: Since $A \cup B = \{-1 \le x \le 2\}$, we have

$$\mathbb{P}(-1 \le x \le 2) = \mathbb{P}\left(\frac{-1-1}{2} \le z \le \frac{2-1}{2}\right) = \mathbb{P}(-1 \le z \le .5) = .6915 - .1587 = .5328$$

(c) (5 points) Find $\mathbb{P}(x \leq 1)$

Solution: Since x has mean 1, this is the area under the lower half of the curve, so $\mathbb{P}(x \leq 1) = \frac{1}{2}$.

Problem 4 - (20 total points) Rumor has it that 10% of Cornell undergrads are believed to be aliens from the planet Krypton. A study is conducted on 400 students using a glowing green crystal of unknown origin to indicate whether or not a given student is an alien. Let x be the number of super-students (aliens) found in such a survey.

(a) (5 points) What is the mean and standard deviation of the binomial random variable x?

Solution: Mean np = 400(.10) = 40, standard deviation $\sqrt{np(1-p)} = \sqrt{40(\frac{9}{10})} = \sqrt{4(9)} = \sqrt{36} = 6$.

(b) (10 points) Find $\mathbb{P}(28 \le x \le 40)$ using the normal approximation to the binomial distribution.

Solution: Using $\mu = 40$ and $\sigma = 6$ from above, we get that

$$\mathbb{P}(28 \le x \le 40) = \mathbb{P}\left(\frac{28 - 40}{6} \le z \le \frac{40 - 40}{6}\right) = \mathbb{P}(-2 \le z \le 0) = \frac{1}{2} - .0228 = .4472$$

(c) (5 points) In a binomial experiment, we can use the expected value formula $\mathbb{E}(x) = np$ to estimate p, the probability of a success. If we assume the expected number of aliens is $\mathbb{E}(x) = 20$, then calculate a better estimate of p (the probability a student is an alien).

Solution: $\mathbb{E}(x) = np = 20$, therefore since n = 400 we solve 400p = 20 and get p = 0.05. (Therefore if the survey found an average of 20 aliens in each group of 400 students instead of the expected 40, we might conclude that there are probably only 5% aliens, not the 10% claimed in the rumor.)

Problem 5 - (20 total points) In a small high school lunch room, there are three groups of tables arranged in a triangle (call them A, B and C). A food-fight happens one day, and during each second of the food-fight, food travels between groups as follows: 50% of the food in A stays in A and the rest goes to B; 50% of the food in B goes to A and the rest goes to C; and 50% of the food in C goes to A and the rest stays in C.

(a) **(3 points)** Draw a transition diagram for this process. Include all of the non-zero probabilities in the sketch.



(b) (3 points) Write down a transition matrix (P) for this process.

Solution: Representing A, B, C with rows 1, 2, and 3 respectively gives

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

(c) (4 points) Find the square of the transition matrix, and circle the position in the final matrix that contains the probability of starting in *B* and ending in A after 2 seconds.

Solution: Squaring P we get that

$$P^2 = \begin{bmatrix} .5 & .25 & .25 \\ (.5) & .25 & .25 \\ .5 & .25 & .25 \end{bmatrix}$$

(d) (4 points) Note P is a *regular*. Use the probability vector $V = [v_1, v_2, v_3]$ and the system of equations VP = V to find the equilibrium (or "long term") distribution of food among groups.

Solution: Multiplying VP and setting it equal to $V = [v_1 \ v_2 \ v_3]$. gives

$$\left[\frac{1}{2}(v_1 + v_2 + v_3) = v_1, \frac{1}{2}v_1 = v_2, \frac{1}{2}(v_2 + v_3) = v_3\right]$$

and additionally we know $v_1 + v_2 + v_3 = 1$ because V is a probability vector.

Solution 1 - "Simple" Shortcuts: Using this last equation and equation 1 above, we substitute in the 1 and see that $v_1 = \frac{1}{2}$. By the second equation above $(\frac{1}{2}v_1 = v_2)$, it therefore follows that $v_2 = \frac{1}{4}$. Using the fact that V is a probability vector agian, we see that $\frac{1}{2} + \frac{1}{4} + v_3 = 1$ thus we have that $v_3 = \frac{1}{4}$. So after thinking about our equations and looking for easy substitutions we have the equilibium distribution

$$V = [.5, .25, .25]$$

Solution 2 - The Sledge Hammer: After we get our equations, our other option is to methodically solve them using some linear algebra tricks. Namely, either Gauss-Jordan or the Echelon method. To begin, we first take either all 4 equations, or we thoughtfully choose 3 equations (hint, use all but 1 from the VP = V equations and the $v_1 + ... + v_3 = 1$ equation.) and write the equations in the proper form (variables all to 1 side, etc.) to make a coefficient matrix:

$$v_1 + v_2 + v_3 = 1$$
$$\frac{1}{2}v_1 - v_2 = 0$$
$$\frac{1}{2}v_2 - \frac{1}{2}v_3 = 0$$

Note we left out the "VP" equation with the most terms. This starts us off with as many zeros in our matrix as possible! We can now write the matrix and do row operations to work up each column from left to right making all the non-diagonal terms zero. (Follow along with scratch paper for easy practice!)

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ \frac{1}{2} & -1 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix} \begin{array}{c} R_1 - 2R_2 \rightarrow R_2 \\ 2R_3 \rightarrow R_3 \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{array}{c} R_1 - R_3 \rightarrow R_1 \\ R_2 - 3R_3 \rightarrow R_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 3 & 1 & 1 \\ R_2 - 3R_3 \rightarrow R_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 4 & 1 \end{bmatrix}$$
$$\begin{array}{c} R_1 - \frac{1}{2}R_3 \rightarrow R_1 \\ R_2 - \frac{1}{4}R_3 \rightarrow R_2 \\ \frac{1}{4}R_3 \rightarrow R_3 \end{array} \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 3 & 0 & \frac{3}{4} \\ 0 & 0 & 1 & \frac{1}{4} \end{bmatrix} \frac{1}{3}R_2 \rightarrow R_2 \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{4} \end{bmatrix}$$

Which gives V = [.5, .25, .25]. Be paranoid and methodical - we don't want mistakes! Watch those negative signs, "easy algebra", and the numbers in the far right column! Double check *each step* before proceeding! Lastly, it can save time to invest a minute or two and recognize how to simplify without using matrices, as in *solution 1* above (e.g. can you easily reduce your equations to fewer variables?). Remember: A little forethough can speed things up and prevent common mistakes.

- (e) Suppose a muffin starts in group A. State whether or not each sequence of states is a possible route for the muffin to follow:
 - (i) (2 points) $A \to B \to C \to A$ Possible. All transitions have non-zero probability.
 - (ii) (2 points) $A \to B \to B \to A$ Not possible. $\mathbb{P}(B \to B) = 0$.
 - (iii) (2 points) $A \to B \to C \to C$ Possible.

Problem 6 - (15 total points) For the matrices given below, determine whether or not they are transition matrices for a Markov chain (as defined in this course). If the matrix is a Markov chain transition matrix, determine whether or not it is regular. (Show your work for full credit.)

(a) **(5 points)** $P_1 = \begin{bmatrix} 1.0 & 0.0 \\ 0.2 & 0.8 \end{bmatrix}$.

Solution: To be a transition matrix, the rows each must sum to 1, the matrix must have entries only between 0 and 1 (including 0 and 1), and it must be square. As such, P_1 is a transition matrix (see pg. 509).

To determine if it is regular or not, we must show that for some number k = 1, 2, 3..., that P^k contains no zero entries (regular), or alternatively that for all such k, P^k always contains at least 1 zero entry ("irregular"). Note the latter case can be shown using the fact that if P^k and P^{k+1} each have a zero at the same location, this zero will exist for all higher powers $\geq k$. Starting with $P_1^2 = \begin{bmatrix} 1.0 & 0.0 \\ 0.36 & 0.64 \end{bmatrix}$, we see the top right entry is 0 in both cases, hence it will be zero for all P^k so by definition, P is not regular (pg. 519).

(NOTE: Showing P does or does not have a *unique* equilibrium vector is only helpful in certain cases, and requires a little more work than we went into in class. We'd be glad to explain to you when this approach is valid, but for our purposes the above approach is the most practical one to determine regular vs. irregular.)

(b) **(5 points)** $P_2 = \begin{bmatrix} 0.0 & 1.0 \\ 0.7 & 0.3 \end{bmatrix}$.

Solution: According to the above definition, P_2 is a transition matrix (pg. 509). Here we see that $P_2^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.21 & 0.79 \end{bmatrix}$ has no zero entries, therefore by definition P_2 is *regular*.

(c) **(5 points)** $P_3 = \begin{bmatrix} 0.5 & 0.3 \\ 0.7 & 0.5 \end{bmatrix}$.

Solution: Note the rows do not sum to 1, therefore this is not a transition matrix according to our definition.