Discussion 2

I. Review

1. Random Varable:
   (a) A random variable is a rule that attaches a numerical value to a chance outcome.
   (b) Discrete Variable: Random variables with a countably infinite set of possible values (i.e. numbers of football games a UW madison student watch every year, how many children are there within a family; think about left-handness, whether or not it's a discrete variable)
   (c) Continuous Variable: Random variables with a continuum of possible values (i.e. heights of female students in UW madison)

2. Binomial Distribution:
   (a) Definiton: the probability distribution for the number of sucesses in a fixed number of independent trials, when the probability of success is the same in each trial.
   Keywords:(Binary, Independent,N fixed, Same probability)
      i. each of the trial has only two outcomes, whcih are denoted as 'success' or 'failure'
      ii. a sample of n independent trials
      iii. n, the sample size is fixed
      iv. each trail has the same probability of 'success' p
   (b) Probability formula: $X \sim$ Binomial(n,p)

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \;\; k = 0, 1, \ldots, n$$

   where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ the number of patterns to choose k objects from n
   (c) : can be viewed as a kind of avearage with respect to probability
   (d) X is the number of 'sucess' in the n indepedent trials
   (e) Factorial: $n! = n(n-1)\ldots 2 \times 1$

3. Expectation of Random Variable:
   (a) The **Expected Value** of a discrete random variable X is:

$$E[X] = \mu = \sum_{i=1}^{N} x_i P(X = x_i)$$

   $x_i' s$ are the values the random variable assumes with positive probability
   (b) The **Variance**: In general, we have the relationship between variance and expectation:

$$Var[X] = E[X^2] - [EX]^2, SD(X) = \sqrt{Var(X)}$$

   For discrete random variables X,

$$Var[X] = \sum_{i=1}^{N} (x_i - \mu)^2 P(X = x_i)$$

   (c) Expected value and variance of Binomial random variable:

$$\text{Mean } E[X] = np, \text{ and Variance } Var[X] = np(1-p).$$

(d) the two linearity rules of expectation:

i. $E(cX) = cE(X)$

ii. $E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ...E(X_n)$

(e) Rules for variance:

i. $Var(cX) = c2Var(X)$

ii. $Var(X1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(Xn)$ if $X_1, x_2, ..X_n$ are independent.

4. Sampling Distribution:

(a) statistic: a numerical value that can be computed from a sample of data.

(b) estimator: a statistic used to estimate the value of a characteristic of a population.

(c) sampling distribution of a statistic: simply the probability distribution of the statistic when the sample is chosen at random.

(d) sample proportion: $\hat{p} = \frac{X}{n}$

i. The possible values of $\hat{p}$ are $0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1$

ii. $P(\hat{p} = \frac{k}{n}) = P(X = k)$

iii. $E(\hat{p}) = p$

iv. $Var(\hat{p}) = \frac{p(1-p)}{n}$

v. $SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$ (Note: The standard deviation of the sampling distribution of an estimate is called the standard error of the estimate)

II.Practice Problems

1. Calcuate $\binom{10}{4}, \binom{25}{3}$

2. For each part, determine if the random variables is binomial or not. If so, state the values of the parameters n and p. If not, explain which assumption(s) of the binomial distribution are violated

(a) Many couples pannning a new family would prefer to have at least one child of each sex. The probability that a couple's child is a boy is 0.512. Suppose that the sex of child is independent of the sex of former children. A new family plans to have 5 children. Let X_1 be the number of boys

(b) Assume that the child's sex depends on the former child. For example, if the first baby is boy, then the probability that the second child is boy will decrease. A new family plans to have 5 children. Let X_2 be the number of boys

(c) There are four blood types, A, B, AB and O. Assume that they have the same probability. We randomly select 100 people. Let X_3 be the number of type A, X_4 be the number of type B, and X_5 be the number of type AB

(d) For these 4 types of blood, if we are just interested in type O or not type O. Let X_6 be the number of type O

3. Suppose $X \sim$ Binomal(10,0.4), calculate the followings

(a) P(X = 3)

(b) $P(X <= 3)$

(c) $P(X > 4)$

(d) $P(1 \leq X \leq 3)$, and $P(1 \leq X < 3)$

(e) Calcuate mean and standard deviation of X

(f) $P(E(X) - SD(X) \leq X \leq E(X) + SD(X))$

(g) $P(E(X) - 2SD(X) \leq X \leq E(X) + 2SD(X))$

4. Many new drugs have been introduced in the last several decades to bring hypertension under control. Supppose a physician agrees to use a new antihypertensive drug on a trail basis on the first 4 untreated hypertensives she encounters in her practice, before deciding whether to adopt the drug for routine use. Let X = the number of patients out of 4 who are brought under control.Then X is a discrete random variable which takes on the value of 0, 1, 2, 3, 4. Suppose from previous experience with the drug, the drug company expects that for any clinical practice the probability that 0 patients out of 4 will be brought under control is 0.008, 1 patient out of 4 is 0.076, 2 patient out of 4 is 0.265, 3 patients out of 4 is 0.411 and all 4 patients is 0.240.

   (a) Draw the distribution table

   (b) Check that it is a correct distribution

   (c) calcuate E(X), Var(X) and SD(X)

5. Consider two random variables Y_1 and Y_2 where $E(Y_1) = 3$ and $E(Y_2) = -1$. In addition, $E(Y_1^2) = 10$ and $E(Y_2^2) = 2$. Calculate the following.

   (a) $E(3Y_1 - 2Y_2)$

   (b) $E(Y_1 - 2Y_1^2 - Y_2 + 6Y_2^2)$

   (c) $Var(Y_1)$ and $Var(Y_2)$

   (d) If additionally assuming Y_1 and Y_2 are independent, calculate $Var(2Y_1 - 3Y_2)$.

6. In the whole population, 30% of people are left-handed. An investigation of 10 people is implemented by a student to estimate the proportion of left-hand people.

   (a) Find the mean and SE for $\hat{p}$

   (b) Calculate $P(|\hat{p} - p| > 3 * SE(\hat{p}))$

III Solutions of the Practice problems

1. (a) $\binom{10}{4} = \frac{10\times9\times8\times7}{4\times3\times2\times1} = 210$

   (b) $\binom{25}{3} = \frac{25*24*23}{3*2} = 2300$

2. (a) Yes. $n = 5$, $p = 0.512$

   (b) No. violation of independent and constant p

   (c) No. Violation of binary outcome. Actually they are multinomial distribution

   (d) Yes. $n = 100$, $p = 0.25$

3. (a) $P(X = 3) = \binom{10}{3}0.4^3 0.6^7 = 120 * 0.064 * 0.028 = 0.215$

   (b) $P(X = 0) = \binom{10}{0}0.4^0 0.6^{10} = 1 * 1 * 0.006 = 0.006$, $P(X = 1) = \binom{10}{1}0.4^1 0.6^9 = 0.0403$, $P(X = 2) = \binom{10}{2}0.4^2 0.6^8 = 0.1209$

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.006 + 0.0403 + 0.1209 + 0.215 = 0.3822 \end{aligned}$$

(c) $P(X > 4) = 1 - P(X <= 3) = 1 - 0.3822 = 0.6178$

(d)

$$P(1 \leq X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0.0403 + 0.1209 + 0.215 = 0.3762$$

$$P(1 \leq X < 3) = P(X = 1) + P(X = 2) = 0.0403 + 0.1209 = 0.1612$$

(e) $EX = np = 4, var(X) = np(1 - p) = 2.4, SD(X) = \sqrt{var(X)} = 1.55$

(f)

$$\begin{aligned} P(4 - 1.55 \leq X \leq 4 + 1.55) &= P(2.45 \leq X \leq 5.55) \\ &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= 0.215 + 0.2508 + 0.2006 = 0.6664 \end{aligned}$$

(g)

$$\begin{aligned} P(4 - 2 * 1.55 \leq X \leq 4 + 2 * 1.55) &= P(0.9 \leq X \leq 7.1) \\ &= 1 - P(X = 0) - P(X = 8) - P(X = 9) - P(X = 10) \\ &= 1 - 0.006 - 0.0106 - 0.0016 - 0.0001 = 0.9817 \end{aligned}$$

4. (a)

k	0	1	2	3	4
P(X=k)	0.008	0.078	0.265	0.411	0.240

(b) all probability is positive and 0.008+0.078+0.265+0.411+0.240 = 1

(c)

$$E(X) = 0 * 0.008 + 1 * 0.076 + 2 * 0.265 + 3 * 0.411 + 4 * 0.240 = 2.80$$

$$E(X^2) = 0^2 * 0.008 + 1^2 * 0.076 + 2^2 * 0.265 + 3^2 * 0.411 + 4^2 * 0.240 = 8.675$$

$$Var(X) = E(X^2) - (EX)^2 = 8.675 - 2.80^2 = 0.835$$

$$SD(X) = \sqrt{Var(X)} = 0.9138$$

5. (a) $E(3Y_1 - 2Y_2) = 3 * E(Y_1) - 2 * E(Y_2) = 3 * 3 - 2 * (-1) = 11$

(b) $E(Y_1 - 2Y_1^2 - Y_2 + 6Y_2^2) = E(Y_1) - 2 * E(Y_1^2) - E(Y_2) + 6 * E(Y_2^2) = 3 - 2 * 10 - (-1) + 6 * 2 = -4$

(c) $Var(Y_1) = E(Y_1^2) - (E(Y_1))^2 = 10 - 3^2 = 1$ and similarly, $Var(Y_2) = 2 - (-1)^2 = 1$.

(d) If additionally assuming Y_1 and Y_2 are independent,
$Var(2Y_1 - 3Y_2) = 2^2 * Var(Y_1) + (-3)^2 * Var(Y_2) = 4 * 1 + 9 * 1 = 13$.

6. Let X denote the number of people surveyed who are left-handed, then X has a binomial distribution of (10, 0.3)

(a) $E(\hat{p}) = E(\frac{X}{n}) = \frac{E(X)}{n} = 0.3$, $Var(\hat{p}) = \frac{p(1-p)}{n} = 0.021$, $SD(\hat{p}) = \sqrt{Var(\hat{p})} = 0.145$

(b) $P(|\hat{p} - p| > 3 * SE(\hat{p})) = P(|\hat{p} - 0.3| > 4 * 0.145) = P(|\hat{p} - 0.3| > 0.58)$
$= P(\hat{p} > 0.88) + P(\hat{p} < -0.28) = P(\hat{p} > 0.88) = P(X > 8.8) = P(X = 9) + P(X = 10)$
$= \binom{10}{9} 0.4^9 * (1 - 0.4) + \binom{10}{10} 0.4^1 0 = 0.00168$