MATH 3215 Project 1 - Binomial Distribution and Expected Value

David Chan

October 2023

1 General Definitions

Experiment: A procedure that is able to be repeated infinitely and has a set of outcomes that are well-defined
Random Experiment: A procedure that is well-defined and has an observable outcome that cannot be predicted exactly in advance
Sample Space: A collection of possible outcomes/results of a (random) experiment
Random Variable: A function mapping possible outcomes of a sample space to a measurable space, a function that assigns values to the outcomes of an experiment
Random: Although difficult to define precisely, "Randomness...exists when some outcomes occur haphazardly, unpredictably, or by chance" (Stanford University). Outcomes can be intrinsically, chaotically random or merely defined as random due to uncertainty caused by incomplete knowledge of a situation.

2 Bernoulli Trials and the Binomial Distribution

A **Bernoulli trial** is a random experiment where there are only 2 possible outcomes, one corresponding to "success" and the other corresponding to "failure." The probability of success does not change among multiple runs of the experiment.

Concept 1: If we conduct a Bernoulli trials with "success" probability p, then the probability of having **exactly** b successes is represented by the following

$$\binom{a}{b} p^b (1-p)^{a-b}$$

Note that this is essentially the PMF (probability mass function) of the binomial distribution.

Example 1: What is the probability of receiving 7 heads when tossing a fair coin 15 times?
(Solution)

$$\binom{15}{7} (\frac{1}{2})^7 (1-\frac{1}{2})^{15-7} = \binom{15}{7} (\frac{1}{2})^{15} = \frac{6435}{2^{15}} = 0.196380615234$$

We have 7 successes with probability $(\frac{1}{2})^7$ and $15-7$ failures with probability $(1-\frac{1}{2})^{15-7}$. The $\binom{15}{7}$ comes from the fact that the 7 successes can occur at any place within the 15 trials, so there are $\binom{15}{7}$ ways of assigning the 7 successes among the 15 trials.

Example 2: What is the probability of getting a "5" on exactly 6 of the rolls when rolling a die 15 times?
(Solution)

$$\binom{15}{6} (\frac{1}{6})^6 (1-\frac{1}{6})^9 = \binom{15}{6} (\frac{1}{6})^6 (\frac{5}{6})^9 = 0.0207905206369$$

We have 6 successes with probability $(\frac{1}{6})^6$ and $15-6$ failures with probability $(1-\frac{1}{6})^{15-6}$. The $\binom{15}{6}$ comes from the fact that the 6 successes can occur at any place within the 15 trials, so there are $\binom{15}{6}$ ways of assigning the 6 successes among the 15 trials.

Example 3: School A's kickball team plays a series of games against School B's kickball team for a total of 7 matches. The school's team to first win 4 games wins the tournament. What is the probability that School A's kickball team wins the tournament?

(Solution) Let $P(A)$ represent the probability that A wins.

$$P(A) = P(A \text{ wins in 4 games})$$
$$+P(A \text{ wins in 5 games}) + P(A \text{ wins in 6 games}) + P(A \text{ wins in 7 games})$$

$$P(A) = p^4 + \binom{4}{3}p^4(1-p) + \binom{5}{3}p^4(1-p)^2 + \binom{6}{3}p^4(1-p)^3$$

We get the term $\binom{4}{3}p^4(1-p)$ from that fact that the probability that School A wins in 5 games is equal to the probability that School A wins 3 out of the first 4 games multiplied by the probability that School A wins the 5th game given that they've won 3 out of the first 4 games. This logic follows for the other terms.

Example 4: Daniel rolls an octahedral die (a die with 8 faces). He will win the game if a rolls a "1." If he does not roll "1" on his first try, however, he will continue rolling until either (1) he rolls a "1" again (which means he loses) or (2) he rolls what he rolled originally (which means he wins). What is the probability that he wins?

(Solution)

$$\frac{1}{8} + \frac{7}{8}\cdot\frac{1}{8} + \frac{7}{8}\cdot\frac{6}{8}\cdot\frac{1}{8} + \frac{7}{8}\left(\frac{6}{8}\right)^2\frac{1}{8} + ... +$$
$$= \frac{1}{8} + \frac{7}{8}\cdot\frac{1}{8}\left(1 + \frac{6}{8} + \left(\frac{6}{8}\right)^2 + \left(\frac{6}{8}\right)^3 + \left(\frac{6}{8}\right)^4 + ... +\right)$$
$$= \frac{1}{8} + \frac{7}{8}\cdot\frac{1}{8}\left(\frac{1}{1-\frac{6}{8}}\right)$$
$$= \frac{1}{8} + \frac{7}{8}\cdot\frac{1}{8}\cdot 4$$
$$= \frac{1}{8} + \frac{28}{64}$$
$$= \frac{2}{16} + \frac{7}{16}$$
$$= \frac{9}{16}$$

There is a $\frac{1}{8}$ chance that Daniel will roll a "1" on the first try. If he doesn't, then he has a $\frac{7}{8}$ chance to roll "not a 1" on the first roll and a $\frac{1}{8}$ chance to roll what he rolled originally on the second roll. He also has a $\frac{7}{8}$ chance to roll "not a 1" on the first roll, a $\frac{6}{8}$ chance to roll "not a 1 or what he rolled originally" on the second roll, and a $\frac{1}{8}$ chance to roll what he rolled originally on the third roll. Additionally, he has a $\frac{7}{8}$ chance to roll "not a 1" on the first roll, a $\frac{6}{8}$ chance to roll "not a 1 or what he rolled originally" on the second roll, a $\frac{6}{8}$ chance to roll "not a 1 or what he rolled originally" on the third roll, and a $\frac{1}{8}$ chance to roll what he rolled originally on the fourth roll. This pattern continues, creating a geometric series. We use the idea of

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

to simplify the infinite geometric series.

Example 5: The lifetime of a carnival toy in hours is a random variable with PDF (probability density function) given below:

$$f(x) = \begin{cases} 0, & x \le 100 \\ \frac{250}{x^2}, & x > 100 \end{cases}$$

What is the probability that 4 of 10 of these toys will die out after 150 hours of use?
(Solution) The probability that a carnival toy will last less than 150 hours is

$$\int_0^{150} f(x)dx = \int_0^{100} 0dx + \int_{100}^{150} \frac{250}{x^2}dx = 0+250\left[-\frac{1}{x}\right]_{100}^{150} = -250\left[\frac{1}{x}\right]_{100}^{150} = -250(\frac{1}{150}-\frac{1}{100}) = -\frac{250}{150}+\frac{250}{100} = -\frac{5}{3}+\frac{5}{2} = \frac{5}{6}$$

Thus, utilizing the idea of the binomial distribution, the probability that 4 of 10 of these toys will die out after 150 hours of use is

$$\binom{10}{4}(\frac{5}{6})^4(\frac{1}{6})^6$$

3 Applications of Expected Value and Random Variables

Concept 2: Suppose we have a random variable Y that assigns a value to every element in the sample space. The expectation of Y, written as $E(Y)$, can then be given by:

$$Y(x_1)p(x_1) + Y(x_2)p(x_2) + ... + Y(x_n)p(x_n)$$

Concept 3: Given a probability space, let $X_1, X_2, ..., X_n$ be random variables. The following holds:

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

Example 6: Lucas and Nicole are arguing over what restaurant they should go. A mutual friend, Joanne, says that the two of them should play a game to calm their dispute and suggests the following: Lucas tosses a fair coin 15 times. If he gets exactly 7 heads out of those 15 tosses, then he gains 7 points. Otherwise, Nicole gains 2 points. What is the expected value of Lucas's final point count? Has Joanne biased the game in favor of someone? If, so, who?
(Solution) We let $E(X)$ represent Lucas's expected value for total point count. (See Example 1 for the probability of receiving 7 heads when tossing a fair coin 15 times.)

$$E(X) = 7(0.196380615234) - 2(1 - 0.196380615234)$$

$$E(X) = 1.37466430664 - 1.60723876953$$

$$E(X) = -0.23257446289$$

Joanne has biased the game in favor of Nicole because $E(X) < 0$.

Example 7: You have been selected to play on the game show Deal or No Deal. In the game show, you are presented with 26 suitcases each with a different amount of cash. The case amounts appear below in dollars:

.01	1,000
1	5,000
5	10,000
10	25,000
25	50,000
50	75,000
75	100,000
100	200,000
200	300,000
300	400,000
400	500,000
500	750,000
750	1,000,000

Assuming no cases have been opened yet, what is the amount of cash you are expected to take home?
(Solution)

$$E(X) = .01 \cdot \frac{1}{26} + 1 \cdot \frac{1}{26} + 5 \cdot \frac{1}{26} + 10 \cdot \frac{1}{26} + 25 \cdot \frac{1}{26} + 50 \cdot \frac{1}{26} + 75 \cdot \frac{1}{26}$$
$$+100 \cdot \frac{1}{26} + 200 \cdot \frac{1}{26} + 300 \cdot \frac{1}{26} + 400 \cdot \frac{1}{26} + 500 \cdot \frac{1}{26} + 750 \cdot \frac{1}{26}$$
$$+1,000 \cdot \frac{1}{26} + 5,000 \cdot \frac{1}{26} + 10,000 \cdot \frac{1}{26} + 25,000 \cdot \frac{1}{26} + 50,000 \cdot \frac{1}{26} + 75,000 \cdot \frac{1}{26} + 100,000 \cdot \frac{1}{26}$$
$$+200,000 \cdot \frac{1}{26} + 300,000 \cdot \frac{1}{26} + 400,000 \cdot \frac{1}{26} + 500,000 \cdot \frac{1}{26} + 750,000 \cdot \frac{1}{26} + 1,000,000 \cdot \frac{1}{26}$$

$$E(X) = \frac{341841601}{2600} \approx \$131,477.54$$

You are expected to take home $\$131,477.54$.

Example 8: You go to a carnival, where guests are invited to play a game upon entrance. The winner will be declared as the person who has drawn the most marbles in the game at the end of the day. There is a large opaque jar where you have probability μ of drawing a green marble. The game ends when you draw a green marble. On average, how many marbles will you draw?
(Solution)

$$E = \mu + (1 - \mu)(1 + E)$$

Note that if you draw a green marble, the game automatically ends. If the marble isn't green, then you are back to the situation where you started, and it will take about E more moves to finish on average.

$$E = \mu + (1 - \mu)(1 + E)$$
$$E = \mu + 1 + E - \mu - \mu E$$
$$0 = 1 - \mu E$$
$$E = \frac{1}{\mu}$$

Example 9: Assume each person in the universe has a different height. k people randomly take a seat around a large, circular table with k chairs. What is the expected number of people taller than both of the people sitting immediately to the ”left” and ”right” of them?
(Solution) Let X_j represent a Bernoulli random variable with the value of 1 if the person in chair j is taller than the people immediately to the left and right of them. $X = \sum_{j=1}^{k} X_j$ thus represents the total number of people taller than the people immediately to the left and right of them. The expected value is thus

$$E(X) = E(\sum_{j=1}^{k} X_j) = \sum_{j=1}^{k} E(X_j)$$

With 3 arbitrary individuals, the probability that the person in between is the tallest is $\frac{1}{3}$. Thus $E(X_j) = \frac{1}{3}$. In conclusion, the expected number of people taller than both the people to the left and right of them is

$$E(X) = \sum_{j=1}^{k} E(X_j) = \frac{k}{3}$$

Example 10: To attract customers to his movie theater, Barry offers a themed action figure each time you come to his theater, randomly chosen from a set of n different figures. How many times should you expect to come to the movie theater to collect all the different figures (approximately)?
(Solution) Let's start with a more concrete example. Suppose $n = 5$. On your first visit, you will definitely get a new figure, so the expected number of visits is 1. To get your second figure, you have a $4/5$ probability of getting a figure that you didn't have previously. To get your third figure, you have a $3/5$ probability of getting a figure that you didn't have previously. To get your fourth figure, you have a $2/5$ probability

of getting a figure that you didn't have previously. To get your fifth figure, you have a 1/5 probability of getting a figure that you didn't have previously. Thus the average number of visits necessary is

$$1 + \frac{1}{\frac{4}{5}} + \frac{1}{\frac{3}{5}} + \frac{1}{\frac{2}{5}} + \frac{1}{\frac{1}{5}} = 1 + \frac{5}{4} + \frac{5}{3} + \frac{5}{2} + 5$$

We can generalize this for all n:

$$1 + ... + \frac{1}{\frac{n-(n-3)}{n}} + \frac{1}{\frac{n-(n-2)}{n}} + \frac{1}{\frac{n-(n-1)}{n}}$$

$$= 1 + ... + \frac{n}{3} + \frac{n}{2} + n$$

$$= n(1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n})$$

Also note that

$$\ln(n) = \int_1^n \frac{1}{t} dt$$

The average number of visits necessary is approximately $n \ln(n)$.

4 Exercises

1. Suppose we perform a Bernoulli trial where $P(x_1) = p_1, P(x_2) = p_2, P(x_3) = p_3$, meaning that the probability of outcome x_j is p_j. The Bernoulli trial terminates only when x_1 happens (success) or x_2 happens (failure). If x_3 happens, then the trial continues. What is the probability that you get a success? Express this probability (1) in terms of only p_1 and p_3, and (2) in terms of only p_1 and p_2.
2. Traits are often determined by a pair of genes. Assume G represents a dominant gene for a trait and g represents a recessive gene for a trait in a pea plant. Thus, a pea plant with the pair of genes GG is fully dominant for the trait, a pea plant with the pair of genes gg is fully recessive for the trait, and a pea plant with the pair of genes Gg is hybrid for the trait. The phenotypes of the fully dominant and hybrid pea plants are the same, and each child pea plant receives 1 gene from each parent. If hybrid pea plant parents have 4 children, what is the probability that 3 of the 4 have the phenotype of the dominant gene?
3. You are offered the following: Someone says they will pay you $m^4 + 2m^2 + m + 3$ dollars for whatever m you roll on a die, where m represents the number that has been rolled. However, you must pay \$400 to play the game each time. You can continue the game for as long as you like, but should you?
4. A **derangement** is a permutation of the elements in a set where $\alpha(i) \neq i$ for all $i = 1, 2, ..., n$. Consider the following scenario. Joseph and Larry are playing a game. Joseph must pay Larry \$2 to start the game. There are 100 marbles with the values $1, 2, ..., 100$ inside a large, opaque jar. Joseph and Larry get their friend Amy to draw out the 100 marbles one by one without looking, resulting in a permutation of the elements. Joseph gets \$6 if the resulting permutation is a derangement. Is this a fair game? If not, explain why and how we can make this game fair (by adjusting Joseph's payoff)? You may find the following fact useful: Given a positive integer n, the number $!n$ derangements of $[n]$ is

$$!n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)!$$

5. You are on a game show similar to *Millionaire* where you have just correctly answered the \$100,000 question. Now you must consider whether to answer the \$250,000 question. If you answer correctly, you will win the \$250,000 prize winnings. However, if you answer incorrectly, you will win only the \$50,000 prize. At what confidence percentage (probability of guessing correctly) should you guess?