Probability Distributions (Chapter 7)

Know how to find the expected outcome of a discrete random variable, (12, 2b; 08, 3a; 05, 2a; 05b, 2a; 04, 4b; 03b, 5b; M2, 5)

Let the random variable X represent the number of caffeinated drinks that a student drinks on a daily basis. The probability distribution of X is shown in the table below.

X	0	1	2	3	4	5	6
P(X)	0.15	0.30	0.25	0.15	0.10	0.04	0.01

Calculate the expected value (the mean) of X.

$E(X) = \mu_x = 0(0.15) + 1(0.30) + 2(0.25) + 3(0.15) + 4(0.10) + 5(0.04) + 6(0.01) = 1.91$

Know what the law of large numbers is and how to describe in the context of a problem, (05, 2b)

Using a sample of 30 students, there was an average of 1.75. A group plans to take another sample of 1000 students and compute the average from those students. How do you expect this average to compare to that of the smaller sample? Justify your response.

We would expect the average to be larger than 1.75, because by the Law of Large numbers our average should be approaching 1.91 (the expected count). The more observations we take, we should approach 1.91

Know how to find the median, quartiles, and interquartile range by looking at a probability distribution for a discrete random variable, (05, 2c; M7, 12; M7, 24)

Using the data about caffeinated drinks, it is obvious that the minimum is 0 and the maximum is 6. Since the 15th percentile is the last 0 and the 45th percentile is the last 1, we can conclude that the first quartile (25th percentile) is also 1, because it would be all 1's between the 15th and 45th percentile. Using that same logic, we can say that the median (50th percentile) is 2 and the third quartile (75th percentile) is 3.

Know how to find the standard deviation of a discrete random variable, (05b, 2a)

$$\sigma^2(x) = \sum (x_i - \mu_x)^2 p_i = (0 - 1.91)^2(0.15) + (1 - 1.91)^2(0.30) + (2 - 1.91)^2(0.25) + \cdots (6 - 1.91)^2(0.01)$$

If you have two normal distributions that are both normal and you add or subtract them, what is the shape of the new distribution? How do you find the average of the new distribution? How do you find the standard deviation of the new distribution? (08b, 5a; 05b, 2b; M7, 26)

If Mr. Merlo's bowling scores follow N(120, 30), and Mr. Furutani's bowling scores follow N(100, 40). How often should Mr. Merlo beat Mr. Furutani if their scores are independent.

Let's look at the difference between their scores

$\mu_M - \mu_F = 120 - 100 = 20$. Mr. Merlo should win by 20 on average.

$\sigma^2_{M-F} = \sigma^2_M + \mu^2_F \rightarrow \sigma^2_{M-F} = 900 + 1600 \rightarrow \sigma_{M-F} = 50$

So, now we know that the difference in the scores follows $N_{M-F}(20,50)$, because Mr. Merlo should win by 20 on average, but there is a standard deviation of 50, because he isn't always going to win by exactly 50.

$$P(Merlo\ Wins) = P(M - F > 0) = P\left(z > \frac{0 - 20}{50}\right) = P(z > -0.4) = 0.6554$$

Merlo should win 65.54% of the time.

More Probability Distributions

Know how to find the standard deviation when you average two variables, (08, 4c)

Using the information from above, what is the average and standard deviation of Merlo and Furutani's average bowling scores? Essentially, if they were to average ($\frac{M+F}{2}$) their scores every time they played, what would the distribution of those numbers be.

$$\mu_{\frac{M+F}{2}} = \frac{\mu_M + \mu_F}{2} = \frac{120 + 100}{2} = 110$$

$$\sigma^2_{\frac{M+F}{2}} = \sigma^2_{\frac{1}{2}M+\frac{1}{2}F} = \sigma^2_{\frac{1}{2}M} + \sigma^2_{\frac{1}{2}F} = 15^2 + 20^2 \quad \rightarrow \quad \sigma_{\frac{1}{2}M+\frac{1}{2}F} = 25$$

Know how to find the average and standard deviation when you combine distributions and change the units(μ_{2a+3b} or σ_{2a+3b}), (12, 6bc; 05b, 2c)

Know that if you add a constant to every number, the average goes up by that constant, but the standard deviation remains the same. If you multiply a constant by every number, both the average and the standard deviation are multiplied by that number.

Multiplying by a constant multiplies your standard deviation→ $\sigma_{3x} = 3\sigma_x$

Adding a constant doesn't change the standard deviation→ $\sigma_{x+10} = \sigma_x$

If you do both, only the multiplication affects the standard deviation→ $\sigma_{3x+10} = 3\sigma_x$

Chapter 8

A local lottery has been created to help the elementary school in the area. The cost of one ticket is $10, and if your ticket is chosen, then you win $20. The people running the lottery guarantee that each ticket has a 40% chance of winning.

If you believe that you are doing a binomial question, state so immediately in your answer, B(n, p), (11b, 2b; 10b, 3a;09, 2b; 07b, 1b; 06b, 6c; 03, 3c). Know how to solve a basic binomial problem (P(X=1)), (14, 3c; 11b, 2b; 07b, 2b; 06b, 6c; 03, 3c; M2, 32)

Using the lottery situation described above, find the probability of having exactly 7 winning lottery tickets if you purchases 10 total tickets.

State distribution --- B(n,p) → B(10, 0.4)

P(X = 7) = $\binom{10}{7}(0.4)^{7}(0.6)^{3} = 0.0425$

Or

Binompdf(n,p,X) = Binompdf(10, 0.4, 7) = 0.0425 (Note: make sure you note that n =10, p =0.4, and X =7)

Can you set up and solve a binomial problem that is cumulative using a calculator($PX \leq 30$)=binomcdf(n, p, 30)? (2010, 4b;10b, 3c;09, 2b). Know how to solve a binomial problem that is the complement of a cumulative problem ($P(X \geq 3) = 1 - P(X \leq 2) = 1 - binomcdf(n, p, 2)$), (11b, 3c; 06, 3b; 06b, 3b)

Using the lottery situation described above, find the probability that you bought at least 7 winning tickets if you purchased 10 total tickets.

State distribution --- B(n,p) → B(10,0.4)

P($X \geq 7$) = 1 – P(X ≤ 6) = 1 – Binomcdf(10, 0.4, 6) = 0.0548

Where n =10, p = 0.4, and X = 6 (Note: make sure you note what the parts of the computer input represent)

Can you find the expected outcome and standard deviation of a binomial distribution? (10, 4a;10b, 3b)—525*

Using the lottery situation described above, how many winning tickets would you expect to hold if you purchased 10 tickets? What would be the standard deviation of the winning number or tickets if you purchased 10 tickets.

$\mu_x = np = 10(0.4) = 4$ tickets

$\sigma_x = \sqrt{np(1-p)} = \sqrt{10(0.4)(0.6)} = 1.549$ tickets

If people purchased 10 tickets, the average number of winning tickets would be 4 tickets with a standard deviation of 1.549 tickets.

Know the conditions for a binomial problem and when a problem is not binomial, (04, 3a; M7, 7; M7, 11)

So far this year Matt Kemp gets a hit in 40% of his at bats. During the season he will probably have 500 at bats, and he gets a bonus if he has more than 200 hits. If X is the number of hits that he gets, does X have a binomial distribution? Explain why or why not

There are four things necessary to create a binomial distribution

1) Set number of observations
2) Success/Failure
3) Probability is constant
4) Observations are independent

In this case, 1 and 2 have been met. He will have 500 observations in which he either gets a hit (success) or does not get a hit (failure). In this scenario, though, 3 and 4 have not been met. It is probably not true that Matt Kemp's probability of success is always 40%. It will most likely be a lot harder to get a hit if he is facing the best pitcher in the league, or a lot easier to get a hit if he is facing the worst pitcher in the league. At bats are not independent, each at bat is affected by several factors, including pitcher, weather, batter health, time or game, etc. So, using the binomial formula would not be appropriate.

Know how to solve a geometric problem where you need to find the probability of taking more than a certain number of tries (Find the probability of taking more than 5 tries to succed), (11b, 3a)

A local lottery has been created to help the elementary school in the area. The cost of one ticket is $10, and if your ticket is chosen, then you win $20. The people running the lottery guarantee that each ticket has a 40% chance of winning. What is the likelihood that you would have to purchase more than 10 tickets in order to get at least one winning ticket.

State distribution → G(0.4) because you are essentially playing until you win

$P(X > 10) = P(X = 11) + P(X = 12) + P(X = 13) + P(X =14) + \ldots$

Since it is infinite, that would be hard to calculate all those individual probabilities. It's easier to think of it this way: Taking longer than 10 tries implies that you MUST lose your first ten tries.

So, P(losing first ten tries) = $0.6^{10} = 0.00605$

Practice Problems For Chapter 7

1) In a particular game, a fair die is tossed. If the number of spots showing is either a 4 or 5, you win $1; if the number of spots showing is a 6, you win $4; and if the number of spots showing is 1, 2, or 3, you win nothing. Let X be the amount that you win on a single play of the game. What is the expected value of X?
   a. $0
   b. $1
   c. $1.33
   d. $2.50
   e. $4
2) In a particular game, a fair die is tossed. If the number of spots showing is either a 4 or 5, you win $1; if the number of spots showing is a 6, you win $4; and if the number of spots showing is 1, 2, or 3, you win nothing. Let X be the amount that you win on a single play of the game. What is the variance of X?
   a. 1
   b. 1.414
   c. 3/2
   d. 2
   e. 13/6
3) The weight of medium-sized tomatoes selected at random from a bin at the local supermarket is a random normal distribution with mean $\mu = 10$ ounces and standard deviation $\sigma = 1$ ounce. Suppose we pick two tomatoes at random from the bin, so the weights of the tomatoes are independent. What is the probability that the sum of the weights of the two tomatoes exceeds 22 ounces is closest to
   a. 0.017
   b. 0.068
   c. 0.079
   d. 0.159
   e. 0.921

4) The admissions policy at a certain university requires that incoming students score in the upper 20% on a standardized test. If the mean score on the test is 510 and the standard deviation of the scores is 80, what is the minimum score that students can earn on the test to meet the admissions requirement? Scores on the test are normally distributed and are reported in intervals of 10.
   a. 520
   b. 580
   c. 590
   d. 600
   e. 620
5) Which of the following is true concerning a normal distribution curve
   I) The curve is symmetric about its mean
   II) The curve is always symmetric about 0
   III) The area under the curve lying within one standard deviation of the mean contains approximately 50% of the scores
      a. I only
      b. II only
      c. III only
      d. I and III only
      e. I, II, and III
6) The scores on the real estate licensing examination given in a particular state are normally distributed with a standard deviation of 70. What is the mean test score if 25% of the applicants score above a 475?
   a. 416
   b. 428
   c. 468
   d. 522
   e. There is not enough information

7) A normal population has a mean of 100 and standard deviation of 16. If a sample of size 64 is drawn from the population, what proportion of the sample should be between 91 and 103?
   a. 0.2875
   b. 0.4987
   c. 0.9050
   d. 0.9319
   e. 0.9332
8) The mean score on a nationally used standardized reading exam was 75 with a standard deviation of four. What would be the z-score for a child who scored an 81?
   a. -1.50
   b. -1.03
   c. 1.03
   d. 1.05
   e. 1.50
9) Evan and Amy are two seniors in a local high school. The grades on a Calculus final which Evan took had an average of 75 and a standard deviation of 7. The grades on a Psychology final which Amy took had an average of 85 with a standard deviation of 4. Even scored an 89 on the Calculus exam and Amy scored a 91 on the Psychology exam. Relative to the other students in their classes, who did better?
   a. Evan
   b. Amy
   c. Both were the same
   d. You cannot compare because we don't know if the distributions were normal
   e. There is not enough information

Solutions

1) B. E(X) = 0(3/6) + 1(2/6) + 4(1/6) = 6/6 = 1
2) D. Using the handy dandy notebook, we can find that the variance of a probability distribution is Var(X) = $(0-1)^2\left(\frac{3}{6}\right) + (1-1)^2\left(\frac{2}{6}\right) + (4-1)^2\left(\frac{1}{6}\right) = \frac{12}{6} = 2$. If you wanted to find the standard deviation, you would square root the variance, $\sqrt{2}$.
3) C. We can think of choosing two tomatoes as two distributions. In other words, the first event is picking a tomato from a N(10,1) distribution, and the second event is picking a tomato from an equal N(10,1) distribution. In that case
   $\mu_1 + \mu_2 = 10 + 10 = 20$. If we choose two tomatoes, we expect the sum to be 20 ounces on average.
   $\sigma_{1+2}^2 = \sigma_1^2 + \mu_2^2 \rightarrow \sigma_{1+2}^2 = 1 + 1 \rightarrow \sigma_{1+2} = \sqrt{2}$. The sum should be twenty give or take 1.414 ounces.

   Now, we have a new distribution. If everybody in the world chose two of these tomatoes and added their weights together, the distribution would be N(20,1.414).

$$P(Sum > 22) = P\left(z > \frac{22-20}{1.414}\right) = P(z > 1.414) = 0.0786$$

4) B. N(510, 80). The top 20% has 80% below it, which produces a z = 0.84 in a normal distribution. $0.84 = \frac{x-510}{80}$. This produces x = 577.2. 570 won't put you in the top 20% but 580 will. You could also use invNorm(0.8, 510, 80) = 577.3 on your calculator to do percentiles.
5) A. I—A normal curve always has the average at the middle. II-The curve is always symmetric around a z = 0, but that doesn't mean that the average is always 0. III-By the Empirical Rule (68-95-99.7 Rule) 68% of the data lies within one standard deviation of the average, not 50%.
6) B. N(μ, 70). We don't know the average, but we do know that the top 25% has a z = 0.68. So, $0.68 = \frac{475-\mu}{70}$. $\mu = 427.4$
7) A. First, we do not need to divide by $\sqrt{n}$, because we are not asking for the likelihood of the 64 AVERAGING a certain amount. In this case it doesn't matter if we pull 64 or 64,000. The proportion that fulfill our qualifications will be the same. So, N(100,16). P(91 < X < 103) = P(-0.56 < z < 0.19) = 0.5714 – 0.2877 = 0.2837. You can also use normalcdf(91,103,100,16) = 0.2875
8) E. $z = \frac{81-75}{4} = 1.50$

A. Evan → $z = \frac{89-75}{7} = 2.00$. Amy → $z = \frac{91-85}{4} = 1.50$. Since Evan was two standard deviations above average and Amy was only 1.50 standard deviations above average, Evan did butter even though his score was lower

Multiple Choice Questions For Chapter 8

1) A survey reveals that 60% of patients who receive prescription medications from their doctors prefer the brand name drug as opposed to the generic brand. In a random sample of 40 patients, what is the probability that at least 25 of them prefer the brand name?
   a) 0.1255
   b) 0.3745
   c) 0.4402
   d) 0.6250
   e) 0.9500

2) A survey indicates that approximately 40% of college students own a cellular phone. In a random sample of 15 college students, what is the probability that at least five own a cell phone?
   a) 0.010
   b) 0.186
   c) 0.217
   d) 0.333
   e) 0.783

3) A report from the Main Department of Inland Fisheries and Wildlife indicates that there occurs on average 1 fatality per 100 collisions between cars and deer. In 300 collisions between a car and a deer, what is the expected number of fatalities and the standard deviation?
   a) Mean: 0.33 standard deviation: 0.01
   b) Mean: 1 standard deviation: 0.01
   c) Mean: 3 standard deviation: 0.172
   d) Mean: 3 standard deviation: 2.97
   e) Mean: 30 standard deviation: 3.0

4) Which of the following are conditions for a binomial experiment?
      i. Number of trials n is a fixed number
      ii. The n trials are independent
      iii. The probability of success p is equal to the probability of failure q
   a) i. only
   b) ii. only
   c) iii only
   d) i. and ii. only
   e) i. ii. and iii.

5) The all-time leader in career batting average among major league baseball players is Ty Cobb with career average of 0.366. This means he got a hit in 36.6% of his official at-bats. What was Cobb's probability of getting at least one hit in 4 official at-bats?
   a) 0.092
   b) 0.134
   c) 0.162
   d) 0.366
   e) 0.838

6) Mr. and Mrs. Jones have 5 children. The 5 births were independent with no multiple births. Assuming that the chances of having a boy or a girl are the same, what is the probability that the family has at least 2 girls?
   a) 0.1875
   b) 0.3125
   c) 0.4000
   d) 0.5000
   e) 0.8125

7) According to a CBS/ New York Times poll taken in 1992, 15% of the public have responded to a telephone call-in poll. In a random group of five people, what is the probability that exactly two have responded to a call-in poll?
A. 0.138
B. 0.165
C. 0.300
D. 0.835
E. 0.973

8) In a 1974 "Dear Abby" letter a woman lamented that she had just given birth to her 8th child, and all were girls! Her doctor had assured her that the chance of the 8th child being a girl was only 1 in 100. What was the real probability that the 8th child would be a girl?
A. 0.01
B. 0.5
C. (0.5)^7
D. (0.5)^8
E. (0.5)^7+(0.5)^8

9) The yearly mortality rate for American men from prostate cancer has been constant for decades at about 25 of every 100,000 men. In a group of 100 American men, what is the probability that at least 1 will die from prostate cancer in a given year?

A. 0.00025
B. 0.0247
C. 0.025
D. 0.9753
E. 0.99975

10) 65% of all divorce cases cite incompatibility as the underlying reason. If four couples file for a divorce, what is the probability that exactly 2 will state incompatibility as the reason?
A. 0.104
B. 0.207
C. 0.254
D. 0.311
E. 0.423

Answers:

1) c) 0.4402
$P(X \geq 25) = 1 - \text{binomcdf}(40, .6, 24) = 1 - .5598 = 0.4402$

2) e) 0.783
n= 15 p= .4 x≥ 5

1- (Binomcdf(15, .4, 4)
1- (.2173)
= 0.7827

3) c)
(300) (1/100) = 300 (0.01) = 3
Expected = 3
Sd= √np(1-p)
Sd= √(330(0.01))(1- 0.01)
Sd= 1.72

4) d) i and ii only
i and ii are both true but iii states the probability of success *p* is equal to the probability of failure *q* which is wrong because the probability of failure would equal 1-p.

5) e) 0.838
n= 4 p=.366 x= 0
1- (binomcdf(4, .366, 0))
= 0.838

6) e) 0.8125
n= 5 p= .5 x=1
1- (binomcdf (5, .5, 1))
1- (.1875)
= 0.8125

7) A: binompdf(5,0.15,2)= 0.138. 5 is the sample taken. 0.15 is the proportion. 2 is the x

8) B: the probability of the next child being a girl is independent of the gender of the previous children.

9) B: 1-(0.99975)^100 = 0.0247

10) Binompdf(4, 0.65, 2)=0.311