Duffy • Statistics & Research Methods • Fall 2024

Test on Thursday, November 21

#### **CHAPTER FOUR: DISCRETE PROBABILITY DISTRIBUTIONS**

A probability distribution shows the likelihood of each possible outcome. This chapter deals with discrete probability distributions, in which there are a set number of possible outcomes in any given range (e.g., no cars, 1 car, 2 cars), as opposed to continuous probability distributions, in which there are infinitely many possible outcomes which can be sorted into discrete categories (e.g., 0 to 10 cm, 10 to 20 cm, 20 to 30 cm). A particularly important discrete probability distribution is the binomial probability distribution, in which a trial such as predicting a coin flip is attempted n times and P(r) represents the probability of exactly r of those n trials being successful.

4-A	Introduction to Probability Distributions	
-----	---	--

discrete variable • continuous variable • probability distribution

- Classify a variable as discrete or continuous.
- **2** Determine whether or not a discrete variable will be treated as continuous.
- **③** Give the probability distribution of an event.
- **4** Calculate the mean and standard deviation of a discrete probability distribution.

### 4-B Geometric and Binomial Probabilities

geometric probability distribution • binomial probability distribution • binomial experiment

- Calculate the probability that the first success will be on the *n*<sup>th</sup> trial.
- **2** Calculate the probability that the first success will be after the *n*<sup>th</sup> trial.
- **③** Calculate the probability of getting exactly *r* successes in a binomial experiment.
- **④** Explain the components of a binomial experiment calculation.
- **6** Calculate the probability of getting at most or at least *r* successes in a binomial experiment.

## 4-C Binomial Distributions

- **1** Calculate binomial probabilities with the calculator.
- **2** Make a histogram for a binomial probability distribution.
- Calculate the mean and standard deviation for number of successes in a binomial distribution.

# Tuesday • 11/5

Thursday • 11/7

### Thursday • 11/14

## 4-A Introduction to Probability Distributions

- A CONTINUOUS Variable has infinitely many possible values within any given range: There is always a possible value between any two other values. Continuous data must be rounded.
- A DISCRETE Variable has a specific number of possible values within any given range: Either it is nonnumerical, or there exist values that it could not be that are between values that it could be. Discrete data are exact.
- Continuous variables are sorted into ranges. Discrete variables with many possible values may also be sorted into ranges and thus treated as continuous variables.
- **1** Classify a variable as discrete or continuous.
  - 1. The variable is continuous unless the data are nonnumerical or there exists an impossible value between two possible values.
  - 2. In general, data that are counted are discrete and data that are measured are continuous.
  - 0
  - a) Shoe size is discrete because, for example, 9 and 9.5 are possible but 9.1 is not.
  - b) Foot length is continuous because it is measured and there are infinitely many possible values between any two given values, such as decimal values between 9 inches and 10 inches.
- **2** Determine whether or not a discrete variable will be treated as continuous.
  - 1. The variable is truly discrete if it is useful to identify the probability of each individual possible value.
    - The variable is treated as continuous if there are a lot of possible values and it is more useful to identify the probability of ranges rather than specific values.
  - 2. In a histogram, labeling the boundaries of the bars as category boundaries (as in chapter 2) is for data treated as continuous, and labeling each bar itself with a single value (as in this chapter) is for truly discrete data.

#### 2

- a) Number of people in a car is truly discrete data. There would be a bar for 1 person, a bar for 2 people, etc.
- b) Number of people on a train would be treated as continuous data. There might be a bar for 0 to 49 people, a bar for 50 to 99 people, etc.
- A PROBABILITY DISTRIBUTION shows all the possible outcomes or ranges of outcomes of an event and how likely each one is. The sum of the probabilities in a probability distribution is 1 (that is, 100%) because it includes all possibilities.

**③** Give the probability distribution of an event.

- 1. List each possible outcome.
- 2. State the probability of each.

**③** Show the probability distribution for a coin flip.

- 1. heads, tails
- 2. 50% chance of heads, 50% chance of tails

The mean of a probability distribution is the same as its expected value:  $\mu = \sum x P(x)$ .

The standard deviation is of a probability distribution is  $\sigma = \sqrt{\Sigma (P(x)(x - \mu)^2)}$ .

- **4** Calculate the mean and standard deviation of a discrete probability distribution.
  - 1. Identify the value *x* of each possible outcome.
  - 2. Identify the probability P(x) of each possible outcome.  $\Sigma P(x)$  must equal 1.
  - 3. Calculate *xP*(*x*) for each possible outcome.
  - 4. To get the mean, calculate  $\mu = \sum x P(x)$ .
  - 5. Subtract  $\mu$  from each x value.
  - 6. Square each difference in step 5.
  - 7. Multiply each square in step 6 by P(x) for that value.
  - 8. To get the variance, add the products in step 7.
  - 9. To get the standard deviation, take the square root of the variance in step 8.

4	A spinner l	has 50 spaces.	The gol	d one is wort	h 50 points,	the 3 rea	d ones are eac	h worth 2	20 points, a	nd the	9 browr	n ones are eac	h worth	h 10 points.
---	-------------	----------------	---------	---------------	--------------	-----------	----------------	-----------	--------------	--------	---------	----------------	---------	--------------

<u>Color</u>	<u>X</u>	<u>P(x)</u>	<u>xP(x)</u>	<u>x — µ</u>	<u>(х —µ)²</u>	$P(x) (x - \mu)^2$
gold	50	$\frac{1}{50}$	1.00	46	2116	42.32
red	20	$\frac{3}{50}$	1.20	16	256	15.36
brown	10	$\frac{9}{50}$	1.80	6	36	6.48
other	0	<u>37</u> 50	<u>0.00</u>	-4	16	<u>11.84</u>
		$\Sigma P(x) = 1.00$	$\mu = 4.00$			$\sigma^2 = 76.00$
						$\sigma \approx 8.72$

## 4-B Geometric and Binomial Probabilities

In the probability distributions below, *n* represents the number of trials, *r* represents the number of successes, *p* represents the probability of success on each trial (which

must be the same for each trial), and *q* represents the probability of failure on each trial. *p* and *q* are complements.

The GEOMETRIC Probability Distribution,  $P(n) = q^{n-1}p$ , gives the probability that the first success will be the  $n^{\text{th}}$  trial.

Similarly,  $P(n) = q^n$  gives the probability that the first success will be after the  $n^{\text{th}}$  trial.

• Calculate the probability that the first success will be after the *n*<sup>th</sup> trial.

- 1. Identify *n*, the number of trials.
- 2. Identify *q*, the probability of failure on each individual trial.
- 3. Calculate  $P(n) = q^n$ .

• Laurel is predicting rolls on 8-sided dice. Find the probability that her first successful prediction will be after her fourth roll.

- 1. There are n = 4 rolls.
- 2. The probability of not getting an 8 is  $q = \frac{7}{8}$  on each roll.
- 3.  $P(4) = \left(\frac{7}{8}\right)^4 = \frac{2401}{4096}$

**2** Calculate the probability that the first success will be on the *n*<sup>th</sup> trial.

- 1. Identify *n* and *q* (see **1**).
- 2. Identify *p*, the probability of success on each individual trial.
- 3. Calculate  $P(n) = q^{n-1}p$ .

2 Laurel is predicting rolls on 8-sided dice. Find the probability that her first successful prediction will be on her fourth roll.

- 1.  $n = 4, q = \frac{7}{8}$
- 2. The probability of a getting an 8 is  $p = \frac{1}{8}$  on each roll.
- 3.  $P(4) = \left(\frac{7}{8}\right)^3 \left(\frac{1}{8}\right) = \frac{343}{4096}$

- The BINOMIAL Probability Distribution,  $P(r) = \binom{n}{r}p^rq^{n\cdot r}$ , gives the probability that exactly r out of n trials of the same independent event will be successes. This type of problem is called a BINOMIAL EXPERIMENT, and it is a special case of 3-C  $\odot$ :  $\binom{n}{r}$  is the number of possible orders, and  $p^rq^{n\cdot r}$  is the probability of each order.
- **③** Calculate the probability of getting exactly *r* successes in a binomial experiment.
  - 1. Identify *n*, the number of trials.
  - 2. Identify *r*, the number of successes.
  - 3. Identify *p*, the probability of success on each individual trial.
  - 4. Identify *q*, the probability of failure on each individual trial.
  - 5. Calculate  $P(r) = \binom{n}{r} p^r q^{n-r}$ .
  - **3** Out of five 6-sided dice, exactly three roll a 6.
  - 1. There are n = 5 rolls.
  - 2. There are r = 3 6's.
  - 3. The probability of a getting a 6 is  $p = \frac{1}{6}$  on each roll.
  - 4. The probability of not getting a 6 is  $q = \frac{5}{6}$  on each roll.
  - 5.  $P(3) = {5 \choose 3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = 10 \left(\frac{1}{216}\right) \left(\frac{25}{36}\right) = \frac{250}{7776}$
- **④** Explain the components of a binomial experiment calculation.
  - 1. *p*′ is the probability of *r* successful trials, each with probability *p* of success (see 3-C **2**).
  - 2.  $q^{n-r}$  is the probability of n r unsuccessful trials, each with probability q of failure (see 3-C **2**).
  - 3.  $\binom{n}{r}$  is the number of possible orders for r of the n trials to be successes and the rest to be failures (see 3-A **1**). Each order has probability  $p^r q^{n-r}$ .
  - 4.  $\binom{n}{r}p^{r}q^{n-r}$  is the probability that one of the  $\binom{n}{r}$  orders will occur, that is, that exactly r of the n trials will be successes (see 3-C 0).
  - Out of five 6-sided dice, exactly three roll a 6.
  - 1.  $\left(\frac{1}{6}\right)^3$  is the probability of three dice all rolling a 6.
  - 2.  $\left(\frac{5}{6}\right)^2$  is the probability of two dice each not rolling a 6.
  - 3.  $\binom{5}{3} = 10$  is the number of possible orders of three 6's and two non-6's.
  - 4.  $\binom{5}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = \frac{250}{7776}$  is the probability that exactly three out of five 6-sided dice roll a 6.
- **O** Calculate the probability of getting at most or at least *r* successes in a binomial experiment.
  - 1. Identify *n*, *p*, and *q* (see **3**).
  - 2. Calculate the probability (see 3-D 3).
  - **(5)** Out of seven 4-sided dice, at least three roll a 4.
  - 1.  $n = 7, p = \frac{1}{4}, q = \frac{3}{4}$
  - 2. The event as stated has five possibilities: P(3) + P(4) + P(5) + P(6) + P(7). The complement has only three: P(0) + P(1) + P(2).
    - $\begin{aligned} P(0) &= \binom{7}{0} \binom{1}{4} \binom{3}{4}^{7} = 1\binom{1}{16384} = \frac{2187}{16384} \\ P(1) &= \binom{7}{1} \binom{1}{4}^{1} \binom{3}{4}^{6} = 1\binom{1}{4} \binom{729}{1096} = \frac{5103}{16384} \\ P(2) &= \binom{7}{2} \binom{1}{4}^{2} \binom{3}{4}^{5} = 1\binom{1}{16} \binom{243}{1024} = \frac{5103}{16384} \\ P(0, 1, \text{ or } 2) &= \frac{2187}{16384} + \frac{5103}{16384} = \frac{12393}{16384} \\ P(\geq 3) &= \frac{16384}{16384} \frac{12393}{16384} = \frac{3991}{16384} \end{aligned}$

#### 4-C Binomial Distributions

Binomial probabilities can be calculated directly on the calculator.

binompdf(n,p,r) is the probability of exactly r successes.

binomcdf(n,p,r) is the probability of at most r successes.

1-binomcdf(n, p, r-1) is the probability of at least r successes, that is, the complement of at most r - 1 successes.

• Calculate binomial probabilities with the calculator.

1. If the scenario is not already written in terms of yielding exactly, at most, or at least r successes, rewrite it.

2. Identify *n*, *r*, and *p*.

3. Plug the variables into the appropriate function above.

• Out of five 6-sided dice, more than two roll a 6.

1. *More than 2* means *at least 3*.

2.  $n = 5, r = 3, p = \frac{1}{6}$ 

3.  $P(3, 4, \text{ or } 5) = 1 - \text{binomcdf}(5, 1/6, 3-1) \approx 3.55\%$ 

A binomial probability distribution can be shown in a histogram, in which each possible value of r has a bar showing its probability.

**2** Make a histogram for a binomial probability distribution.

1. Identify *n*, *p*, and *q*.

2. Calculate the height of each bar by binompdf(n, p, r). n and p are the same each time, but a different r is used for each bar (0 through n).

3. Label the x-axis as the number of whatever is being counted as successes. Call this variable r, and scale the axis from 0 to n.

4. Label the y-axis as P(r), and scale the axis from 0% up to at least as high as the highest bar.

4. Graph each bar.

5. Title the graph.

2 Make a histogram for the probability distribution of predicting 4 rolls on a 4-sided die.

1.  $n = 4, p = \frac{1}{4}, q = \frac{3}{4}$ 

2. P(0) = binompdf(4, 1/4, 0) = .316

P(1) = binompdf(4, 1/4, 1) = .422

P(2) = binompdf(4, 1/4, 2) = .211

P(3) = binompdf(4, 1/4, 3) = .047

P(4) = binompdf(4, 1/4, 4) = .003

The mean and expected value of a binomial distribution is  $\mu = np$ .  $\mu$  is the most likely outcome in the distribution, or it is beetween the two most likely outcomes if  $\mu$ 

itself is not possible (such as 2.5 successes).

The standard deviation of a binomial distribution is  $\sigma = \sqrt{npq}$ .

Calculate the mean and standard deviation for number of successes in a binomial distribution.

1. Identify *n*, *p*, and *q*.

2. Calculate  $\mu = np$ .

3. Calculate  $\sigma = \sqrt{npq}$ .

Sind the mean and standard deviation for number of correct predictions out of 4 rolls on a 4-sided die.

1.  $n = 4, p = \frac{1}{4}, q = \frac{3}{4}$ 

2.  $\mu = 4(\frac{1}{4}) = 1$ 

 $\sigma = \sqrt{4(1/4)(3/4)} \approx .866$