The Binomial Probability Distribution

MATH 130, *Elements of Statistics I*

J Robert Buchanan

Department of Mathematics

Fall 2023

Objectives

After this lesson we will be able to:

- ▶ determine whether a probability experiment is a binomial experiment,
- ▶ compute probabilities of binomial experiments,
- ▶ compute the mean and standard deviation of a binomial random variable,
- ▶ construct binomial probability histograms.

Binomial Experiments

- A **binomial experiment** repeats a simple experiment several times.
- The simple experiment has only two outcomes.
- The binomial experiment counts the number of outcomes of each of the two types.

Example

Flip a coin 10 times and count the number of heads and tails that occur.

Criteria

Theorem (Criteria for a Binomial Probability Experiment)

*An experiment is said to be a **binomial experiment** if*

1. *the experiment is performed a fixed number of times. Each repetition of the experiment is called a **trial**.*
2. *the trials are all independent. The outcome of one trial does not affect the outcome of any other trial.*
3. *for each trial, there are two mutually exclusive outcomes generally thought of as "success" or "failure".*
4. *the probability of success is the same for each trial.*

Notation

- ▶ Let n be the number of independent trials of the experiment.
- ▶ Let p be the probability of success (and $1 - p$ be the probability of failure).
- ▶ Let X be the random variable denoting the number of successes in the n trials of the binomial experiment.

$$0 \leq X \leq n$$

Which if the following situations describe binomial experiments?

1. A test consists of 10 True/False questions and X represents the number of questions answered correctly by guessing.

2. A test consists of 10 multiple choice (5 choices per question) questions and X represents the number of questions answered correctly by guessing.

Which if the following situations describe binomial experiments?

1. An experiment consists of drawing five cards from a well-shuffled deck with replacement. The drawn card is identified as a "heart" or "not a heart". Random variable X represents the number of hearts drawn.

2. An experiment consists of drawing five cards from a well-shuffled deck without replacement. The drawn card is identified as a "heart" or "not a heart". Random variable X represents the number of hearts drawn.

Binomial Probabilities

The probability of x successes out of n trials of a binomial experiment for which the probability of success on a single trial is p is

$$P(x) = (_nC_x)\, p^x (1-p)^{n-x},$$

for $x = 0, 1, \ldots, n$.

Binomial Probabilities

The probability of x successes out of n trials of a binomial experiment for which the probability of success on a single trial is p is

$$P(x) = (_nC_x)\, p^x (1-p)^{n-x},$$

for $x = 0, 1, \ldots, n$.

Example
What is the probability that in 12 flips of a fair coin that exactly 4 heads will result?

Binomial Probabilities

The probability of *x* successes out of *n* trials of a binomial experiment for which the probability of success on a single trial is *p* is

$$P(x) = (_nC_x)\, p^x (1 - p)^{n-x},$$

for $x = 0, 1, \ldots, n$.

Example

What is the probability that in 12 flips of a fair coin that exactly 4 heads will result?

$$P(4) = (_{12}C_4)(0.5)^4(1 - 0.5)^{12-4} = (495)(0.5)^4(0.5)^8 = 0.1208$$

Table III of Appendix A (pages A–3 through A–6) lists pre-computed values of the binomial probability formula.

- ▶ Table III summarizes the cases of $n = 2, 3, \ldots, 12, 15, 20$.
- ▶ The binomial probabilities for $p = 0.01, 0.05, 0.10, \ldots, 0.95$ are listed.

Binomial Probability Tables (1 of 2)

Table III of Appendix A (pages A–3 through A–6) lists pre-computed values of the binomial probability formula.

- ▶ Table III summarizes the cases of $n = 2, 3, \ldots, 12, 15, 20$.
- ▶ The binomial probabilities for $p = 0.01, 0.05, 0.10, \ldots, 0.95$ are listed.

Example

What is the probability that in 12 flips of a fair coin that exactly 7 heads will result?

Binomial Probability Tables (2 of 2)

Table IV of Appendix A (pages A–7 through A–10) lists pre-computed **cumulative** values of the binomial probability formula.

▶ The cumulative value is $P(x \leq m)$,

$$P(x \leq m) = \sum_{i=0}^{m} (_nC_i)p^i(1-p)^{n-i}$$

▶ Table IV summarizes the cases of $n = 2, 3, \ldots, 12, 15, 20$.
▶ The binomial probabilities for $p = 0.01, 0.05, 0.10, \ldots, 0.95$ are listed.

Binomial Probability Tables (2 of 2)

Table IV of Appendix A (pages A–7 through A–10) lists pre-computed **cumulative** values of the binomial probability formula.

▶ The cumulative value is $P(x \leq m)$,

$$P(x \leq m) = \sum_{i=0}^{m} (_nC_i)p^i(1-p)^{n-i}$$

▶ Table IV summarizes the cases of $n = 2, 3, \ldots, 12, 15, 20$.
▶ The binomial probabilities for $p = 0.01, 0.05, 0.10, \ldots, 0.95$ are listed.

Example
What is the probability that in 12 flips of a fair coin that 7 or fewer heads will result?

Example

The manager of a grocery store guarantees that a carton of 12 eggs will contain no more than one "bad egg". If the probability that an individual egg is bad is $p = 0.05$, what is the probability that the manager will have to replace an entire carton?

Example

The manager of a grocery store guarantees that a carton of 12 eggs will contain no more than one "bad egg". If the probability that an individual egg is bad is $p = 0.05$, what is the probability that the manager will have to replace an entire carton?

Let X be the number of "bad eggs" in a carton. A carton must be replaced if $X > 1$.

$$P(x > 1) = 1 - P(x \leq 1) = 1 - 0.8816 = 0.1184.$$

Mean and Standard Deviation

Theorem

A binomial experiment with n independent trials and probability of success p on a trial has a mean and standard deviation given by the formulas:

$$\begin{aligned} \mu_X &= np \\ \sigma_X &= \sqrt{np(1-p)}. \end{aligned}$$

Examples

There is a 90% chance that a pizza from TelePizza will be delivered in less that 30 minutes. If a pizza is not delivered in less than 30 minutes, the next order is free. Suppose TelePizza must process 300 delivery orders per day. What is the mean and standard deviation in the number of pizzas delivered on time?

Examples

There is a 90% chance that a pizza from TelePizza will be delivered in less that 30 minutes. If a pizza is not delivered in less than 30 minutes, the next order is free. Suppose TelePizza must process 300 delivery orders per day. What is the mean and standard deviation in the number of pizzas delivered on time?

$$
\begin{aligned}
\mu_X &= np = (300)(0.90) = 270.0 \\
\sigma_X &= \sqrt{np(1-p)} = \sqrt{(300)(0.90)(1-0.90)} \approx 5.2
\end{aligned}
$$

Squid Game: Glass Bridge
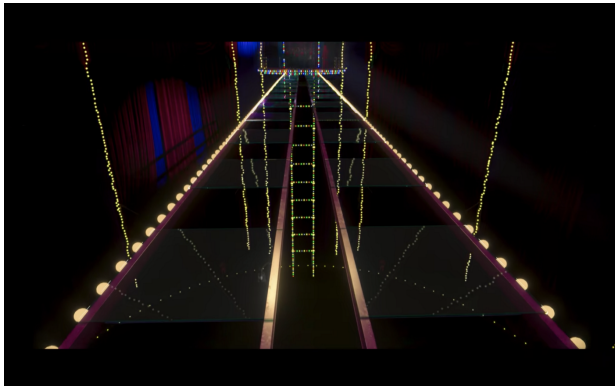
The glass bridge game had two rows of 18 steps each. At each step a player must choose between 2 sheets of glass, one is regular glass which will break and the other is tempered glass which can hold the weight of two people. There are 16 people trying to cross the bridge in a predetermined order.

You Don't Want to be the First Player

- ▶ The first person to play have a 1/2 probability of selecting the tempered glass from a single pair of glass sheets.
- ▶ The probability of the first player guessing all 18 steps correctly is

$$\frac{1}{2^{18}} = \frac{1}{262,144}.$$

- ▶ The first player is not likely to across the entire bridge, but the successful steps they take will provide an advantage to the players which follow.

On Average How Many Steps Does a Player Take?

▶ If a player is facing a string of unknown steps ahead of them, on average who many steps will they take successfully (tempered glass) before they guess incorrectly (regular glass)?

▶ The probabiity that the first incorrect guess is the kth step is

$$P(X = k) = \left(1 - \frac{1}{2}\right)^{k-1} \frac{1}{2} = \frac{1}{2^k}.$$
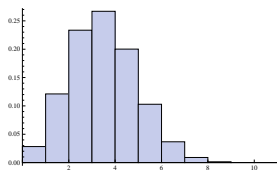
▶ The mean number of correct steps is

$$E(X) = \sum_{k=1}^{18} \frac{k}{2^k} = 1.99992 \approx 2.$$

On average each player will reveal 2 correct steps to the players coming behind them.
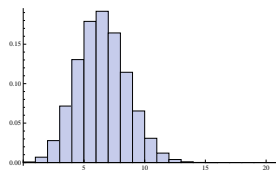
▶ If there are 16 players and 18 steps, then we should expect 9 players to fail and 7 to succeed.
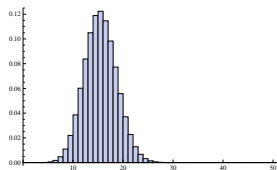
Histograms

Consider the histograms of the binomial probability distribution for $p = 0.30$ and three different values of n.

Observation

As the number of trials n of a binomial experiment increases, the probability distribution of the random variable X becomes bell-shaped. If $np(1 - p) \geq 10$, the probability distribution will be bell-shaped.

Hence when $np(1 - p) \geq 10$ we may use the Empirical Rule to identify unusual observations in a binomial experiment.

Example

There is a 90% chance that a pizza from TelePizza will be delivered in less that 30 minutes. If a pizza is not delivered in less than 30 minutes, the next order is free. Suppose TelePizza must process 300 delivery orders per day.

1. According to the Empirical Rule, between what two values would 95% of the daily on-time deliveries fall?

2. Would it be unusual to find that only 244 pizzas out of 300 were delivered on time?

Example

There is a 90% chance that a pizza from TelePizza will be delivered in less that 30 minutes. If a pizza is not delivered in less than 30 minutes, the next order is free. Suppose TelePizza must process 300 delivery orders per day.

1. According to the Empirical Rule, between what two values would 95% of the daily on-time deliveries fall?

   $(\mu_X - 2\sigma_X, \mu_X + 2\sigma_X) = (270 - (2)(5.2), 270 + (2)(5.2)) = (259.6, 280.4)$

2. Would it be unusual to find that only 244 pizzas out of 300 were delivered on time?

Example

There is a 90% chance that a pizza from TelePizza will be delivered in less that 30 minutes. If a pizza is not delivered in less than 30 minutes, the next order is free. Suppose TelePizza must process 300 delivery orders per day.

1. According to the Empirical Rule, between what two values would 95% of the daily on-time deliveries fall?

   $$(\mu_X - 2\sigma_X, \mu_X + 2\sigma_X) = (270 - (2)(5.2), 270 + (2)(5.2)) = (259.6, 280.4)$$

2. Would it be unusual to find that only 244 pizzas out of 300 were delivered on time?
   Yes, since 244 is 5 standard deviations below the mean.

Example

The National Transportation Safety Board (NTSB) has found that 47% of airline injuries are caused by seat failure. Two hundred cases of airline injuries are selected at random.

1. What is the mean, variance, and standard deviation for the number of injuries caused by seat failure in this group of 200 injuries?
2. According to the Empirical Rule, between what two values would 95% of the injuries due to seat failure fall?
3. Would it be unusual to find that only 105 injures were due to seat failure?

Solution

1. Mean, variance, and standard deviation:

$$
\begin{aligned}
\mu_X &= (200)(0.47) = 94.0 \\
\sigma_X^2 &= (200)(0.47)(1 - 0.47) = 49.8 \\
\sigma_X &= \sqrt{49.8} = 7.1
\end{aligned}
$$

2. According to the Empirical Rule, between what two values would 95% of the injuries due to seat failure fall?
$(\mu_X - 2\sigma_X, \mu_X + 2\sigma_X) = (94 - (2)(7.1), 94 + (2)(7.1)) = (79.8, 108.2)$

3. Would it be unusual to find that only 105 injures were due to seat failure?
Not unusual, since 105 failures is in the middle 95% of the range of the random variable.