The Binomial Distribution

The **binomial distribution** describes how likely certain events are to occur under specific conditions. The binomial distribution is a **discrete** distribution — a binomial variable can only take whole numbers as values.

An experiment can be described by the binomial distribution if it has these conditions:

- the experiment consists of a *fixed* number of *independent* trials
- each trial has only two possible outcomes, known as a **success** and a **failure**
- the probability of a success in every trial is the same

Let's look at two experiments to better understand these conditions. In the first experiment, carnival-goers spin a wheel with 24 spaces, four of which are marked with a star. A player gets three spins. In the second experiment, a person draws from a bag containing five red stones and five green stones until she has three the same colour. Both experiments are a series of **trials**: repeating the same action to see what happens.

Criterion 1: The number of trials must stay the same and each trial must be independent of the others to describe the experiment by the binomial distribution. If you and I do the first experiment, we each get exactly three spins, and the spins don't affect each other. This fits criterion 1. If you and I do the second experiment, we don't know how many draws it will take to get three stones of the same colour. I might do it in three and you might do it in five. This does not fit criterion 1.

Criterion 2: Each trial has only two possible outcomes – success or failure. In the first experiment, a space with a star is considered a success and a space without a star a failure. Since there are 24 spaces on the wheel, there are 24 actual outcomes, but we can group the outcomes into two classes (star or no star), which meets criterion 2. In the second experiment, there are two colours, but the desired outcome of the next draw depends on what you or I have already drawn. If I have two red stones and you have two green stones, we're each hoping for a different colour on our next draw. We do not have two outcomes since we cannot clearly define success and failure by colour.

Criterion 3: The probability of success in each trial is the same. If only spins that make a few revolutions are allowed, the spins are random. The probability of getting a starred section is $\frac{4}{24} = \frac{1}{6}$. The probability of a success is always the same, fulfilling criterion 3. As stones are drawn out of the bag, though, the probability of the next colour drawn changes. At the beginning, P(red) = $\frac{5}{10}$ = 50%. If the first stone is red, this changes to P(red) = $\frac{4}{9}$ = 44%. The probability of each draw changes. In general, any experiment involving drawing without replacement cannot be described by the binomial distribution.

To describe a binomial distribution, we need two **parameters** — numbers that define the distribution: the number of trials, n, and the probability of a success during a trial, p. (The probability of failure is the variable q, but $q = 1 - p$.) We can write that a statistical variable is distributed binomially, and define how in terms of n and p.

If X is a binomially distributed variable, then the probability of x successes, $P(X = x) = {}_nC_x \cdot p^k \cdot q^{n-k}$. Your textbook will tell you that "${}_nC_x$" is $\frac{n!}{x!\,(n-x)!}$, and you can calculate it that way, but you'll be using your calculator to do it, and the calculator will evaluate that entire expression for you. One of the buttons may have "nCr" printed above it, or it may be accessible through a STATS menu, but every scientific calculator will take care of that calculation for you. There's no reason to do it by hand. See your calculator's manual or talk to a tutor to learn how it works.

Example: Consider the carnival game. Find the probability that:
(a) exactly 1 star occurs in 3 spins.
(b) there are either 2 stars or 2 non-stars in 5 spins.
(c) more than 1 star occurs in 20 spins.

Solution: Let X represent the number of stars in each experiment.

(a) Here, n = 3 and p = ⅙, and we want P(X = 1).

$$\begin{aligned} P(X = 1) &= {}_nC_x \cdot p^k \cdot q^{n-k} \\ &= {}_3C_1 \cdot (1/6)^1 \cdot (1 - 1/6)^{3-1} \\ &= 3 \cdot 1/6 \cdot (5/6)^2 \\ &= 25/72 \approx 0.35 \end{aligned}$$

(b) Here, n = 5 and p = ⅙. Getting 2 non-stars means getting 3 stars. We want P(X = 2 or 3). Since the two events in the question are disjoint — we can't get exactly 2 stars and exactly 3 stars at the same time — we can calculate their probabilities separately and add them.

$$\begin{aligned} P(X = 2 \text{ or } 3) &= P(X = 2) + P(X = 3) \\ &= {}_5C_2 \cdot (1/6)^2 \cdot (1 - 1/6)^{5-2} + {}_5C_3 \cdot (1/6)^3 \cdot (1 - 1/6)^{5-3} \\ &= 10 \cdot (1/6)^2 \cdot (5/6)^3 + 10 \cdot (1/6)^3 \cdot (5/6)^2 \\ &= 125/648 \approx 0.19 \end{aligned}$$

(c) Here n = 20 and p = ⅙, and we want P(X > 1). We could do this the same way we did part (b), but that would require calculating all probabilities from 2 up to 20, which is 19 steps! An easier way is to use the reverse of the situation and subtract from 1. If the experiment did not produce more than 1 star, then it produced either 0 stars or 1. This is much easier to work with:

$$\begin{aligned} P(X > 1) &= 1 - P(X \le 1) \\ &= 1 - [P(X = 0) + P(X = 1)] \\ &= 1 - [{}_{20}C_0 \cdot (1/6)^0 \cdot (1 - 1/6)^{20-0} + {}_{20}C_1 \cdot (1/6)^1 \cdot (1 - 1/6)^{20-1}] \\ &= 1 - [1 \cdot 1 \cdot (5/6)^{20} + 20 \cdot 1/6 \cdot (5/6)^{19}] \\ &= 1 - 0.13042\ldots \\ &= 0.86958\ldots \approx 0.87 \end{aligned}$$

The binomial distribution can be used to model the results of surveys. If a certain proportion of the population, p, would say yes to a survey question, then the probability of speaking to someone who would say yes is also p. If you know how many survey respondents you're going to look for, n, then asking each respondent represents one trial in the overall experiment of asking that survey question. It's not *perfectly* binomial, since we don't allow the same person to answer a survey question twice, so we're drawing people without replacement, but with a large enough population to draw from, and with a sample that is no more than 5% of the population, it's quite close.

SIGNIFICANCE

We can also calculate the mean and the spread for the binomial distribution. Over weeks of watching people spin the wheel at the carnival, the person running the game will know about how many stars each person is likely to spin, and how far from this average people are likely to get.

If X is a binomially distributed variable, then $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

For the carnival game, X is the number of stars a player spins, and X is normally distributed. For X, n = 3 and p = ⅙, so μ is 3 · ⅙ = 0.5 and $\sigma = \sqrt{3 \cdot \frac{1}{6} \cdot \frac{5}{6}}$ = 0.6455….

We can use these figures to determine what a significantly high or low result would look like. A result is significantly high if it is μ + 2σ or greater. 0.5 + 2 × 0.6455 = 1.791… but since this distribution only takes on whole number values, getting 2 or 3 stars would be considered significantly high. (The real world implication is that the player who gets 2 or more stars was quite lucky to do so.) A result is significantly low if it is μ − 2σ or less. 0.5 − 2 × 0.6455 = −0.791… but since this distribution only takes on non-negative values, no results would be significantly low.

EXERCISES

A. Determine whether each of these statistical experiments can be modelled by the binomial distribution. If they can, define the statistical variable X and determine the values of n and p for the distribution. If they cannot, state which of the three conditions from the first page on this handout is violated.

1) A tutor in the Learning Centre wants to know whether the way a student faces is an indicator of whether that student will come for help. In one week the tutor tracks one randomly chosen student per day who sits facing the tutoring desks, and finds that over a series of one-week experiments, those students ask for help 60% of the time.

2) You grab five jelly beans out of a jar, hoping to get red one. On the manufacturer's website, it says that 23% of the jelly beans it sells in every package are red.

3) A child learns to hit a baseball by practicing. At the beginning of the day, he can't hit at all, but by the end, he has a batting average of 0.100. An experiment consists of 3 consecutive pitches from this day.

4) A coin is weighted so that it is unfair. It comes up heads 98% of the time rather than 50% of the time. You challenge a friend to a bet saying, "I'll flip this coin, best two out of three. Heads I win, tails I lose."

5) A nurse in a private care facility wants to show that more patients are being admitted than the current staff can care for. She believes the staff can handle six intakes per day. Over a period of two months, she finds that on 0.233 of the days, the facility admits seven or more patients. She goes to her boss with the results for a typical work week of five days.

B. For each of the situations in Part A that were modeled by the binomial distribution, calculate the mean and standard deviation for the distribution, and then determine what values are significantly high and significantly low.

C. For the distributions below, calculate $P(X = x)$.

1) $n = 5$, $p = 0.5$, $x = 2$
2) $n = 3$, $p = 0.75$, $x = 2$
3) $n = 7$, $p = 0.24$, $x = 3$
4) $n = 20$, $p = 0.05$, $x = 1$
5) $n = 10$, $p = 0.33$, $x = 3$
6) $n = 8$, $p = 0.125$, $x = 0$

D. The following experiments follow the binomial distribution. Find the probabilities of the requested outcomes.

1) The carnival game from the main text of this handout. Let X = the number of stars. For X, $n = 3$ and $p = \frac{1}{6}$. Find $P(X = 2)$.

2) At a sleep clinic, they find that results are inconclusive 0.03% of the time due to interference with the equipment used to measure a patient's sleep cycles. The clinic has five sleep rooms that are always booked full. Find the probability that exactly one of the patient's results will be inconclusive on a given night.

3) A quality control technician examines cartons of a dozen eggs for broken eggs. She finds that the probability of finding a broken egg is 0.0075. What is the probability that a given carton has no broken eggs in it?

4) A seasoned darts player throws three darts at the target. He hits the "20" section of the dartboard 72% of the time. Find the probability that he hits the "20" section at least once in 3 darts.

SOLUTIONS

A. (1) yes; X = the number of students who ask for help; $n = 5$, $p = 0.60$ (2) no; there is drawing without replacement, so the probability of success changes. (3) no; the probability changes, and the trials are not independent because the child is learning and improving as he practices. (4) yes; X = the number of heads; $n = 3$, $p = 0.98$. The coin is unfair, but it's *consistently* unfair. (5) yes; X = the number of days with a high number of intakes; $n = 5$, $p = 0.233$

B. (1) $\mu = 3$, $\sigma = 1.1$, nothing significantly high, 0 is significantly low
(4) $\mu = 2.94$; $\sigma = 0.24$; nothing significantly high; 0, 1, 2 are significantly low
(5) $\mu = 1.17$; $\sigma = 0.95$; 4, 5 significantly high; nothing significantly low

C. (1) $P(X = 2) = 0.3125$ (2) $P(X = 2) = 0.4219\ldots$ (3) $P(X = 3) = 0.1614\ldots$
(4) $P(X = 1) = 0.3774\ldots$ (5) $P(X = 3) = 0.2614\ldots$ (6) $P(X = 0) = 0.3436\ldots$

D. (1) $P(X = 2) = 0.0694\ldots$ (2) $n = 5$, $p = 0.0003$, $P(X = 1) = 0.001498\ldots$
(3) $n = 12$, $p = 0.0075$, $P(X = 0) = 0.9136\ldots$
(4) $n = 3$, $p = 0.72$, $P(X \geq 1) = 1 - P(X = 0) = 0.9780\ldots$