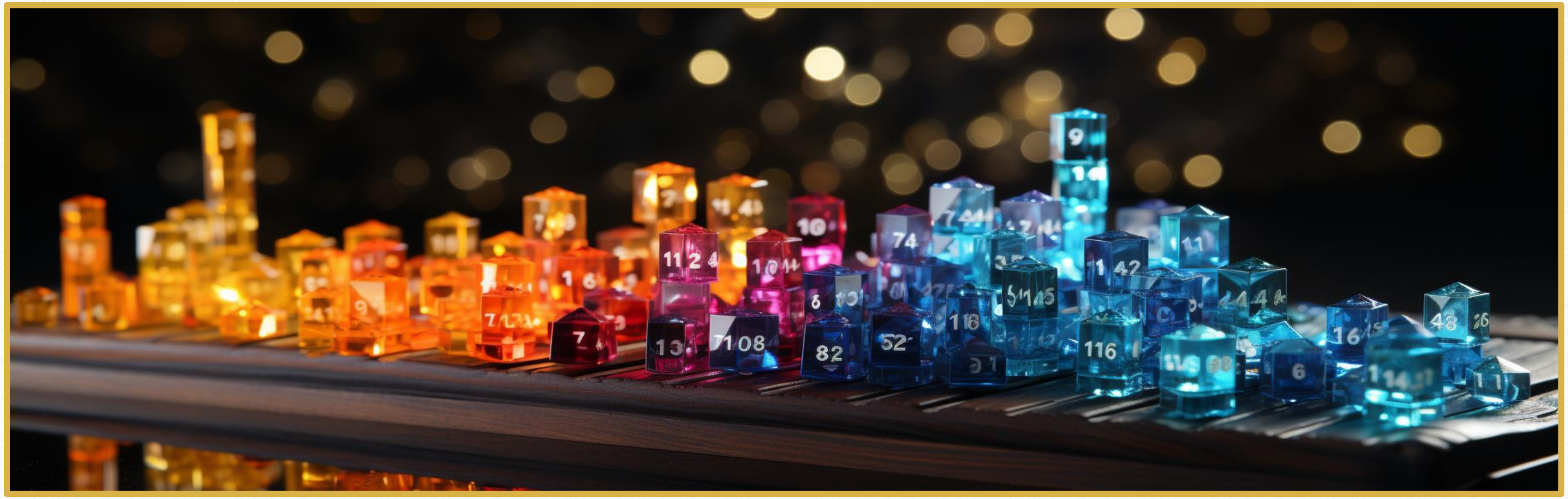
UXG1205 Lecture
9. Binomial Distributions

LIN QINJIE

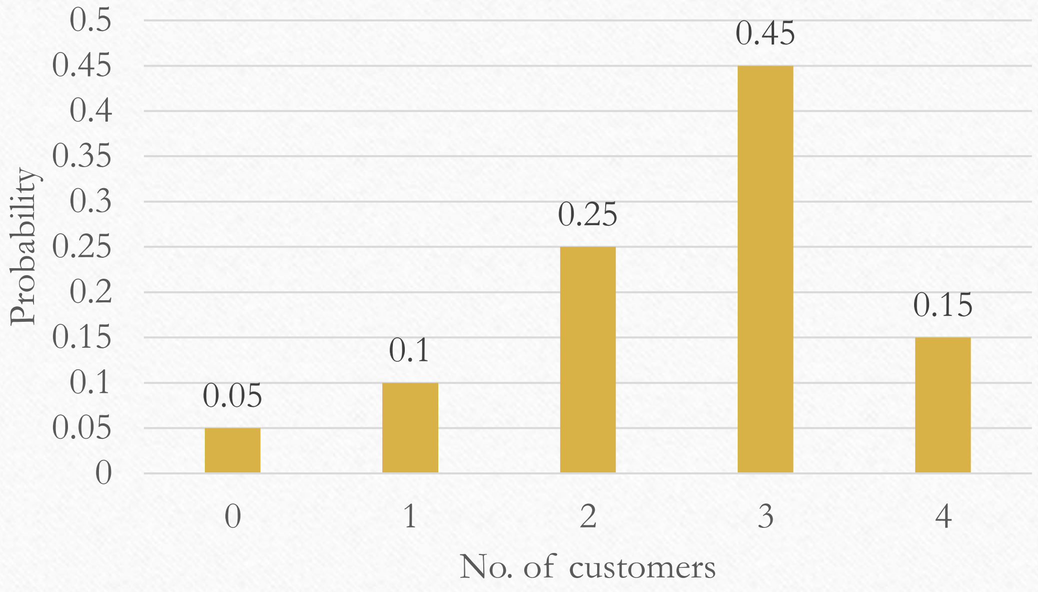
Outline

- Probability Distribution Family
- Binomial Distributions



Probability Distribution Family

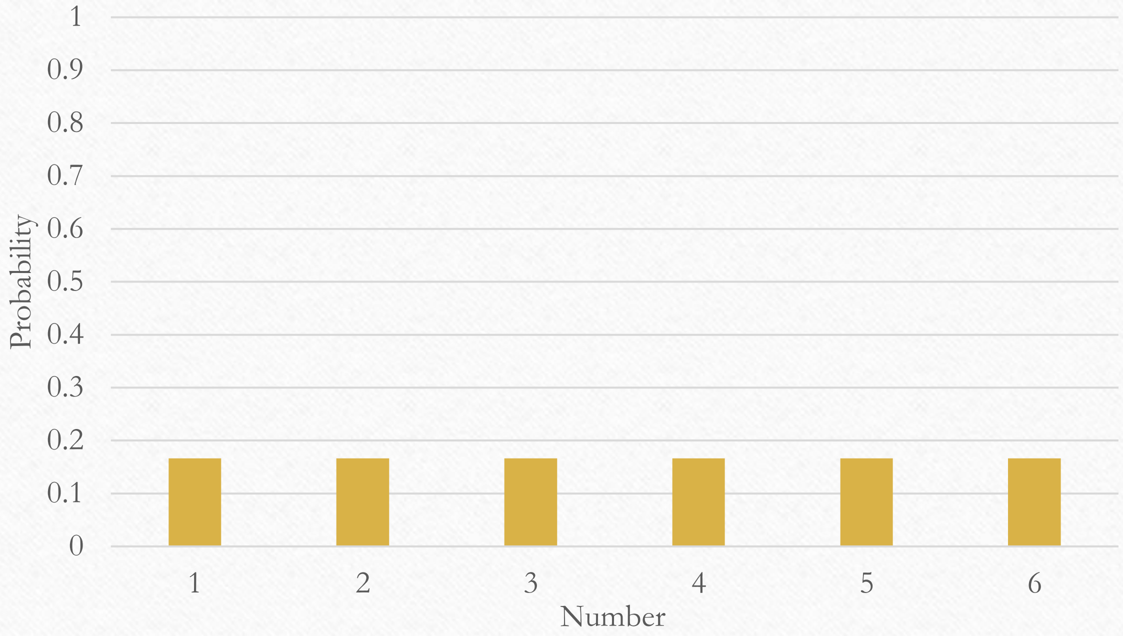
- Definition
- Parameters and outcome of interest



- In the previous topic, Julie constructed a **probability distribution** to show the number of customers that Good Book Store may have in a day.
- How did she get these values?

- To get those probabilities, it is likely that she undertook extensive data collection.

- One approach to this is by gathering observations.
  - Leveraging the data acquired from a specific sample period or population allows us to construct a probability distribution, providing insights into potential future events.



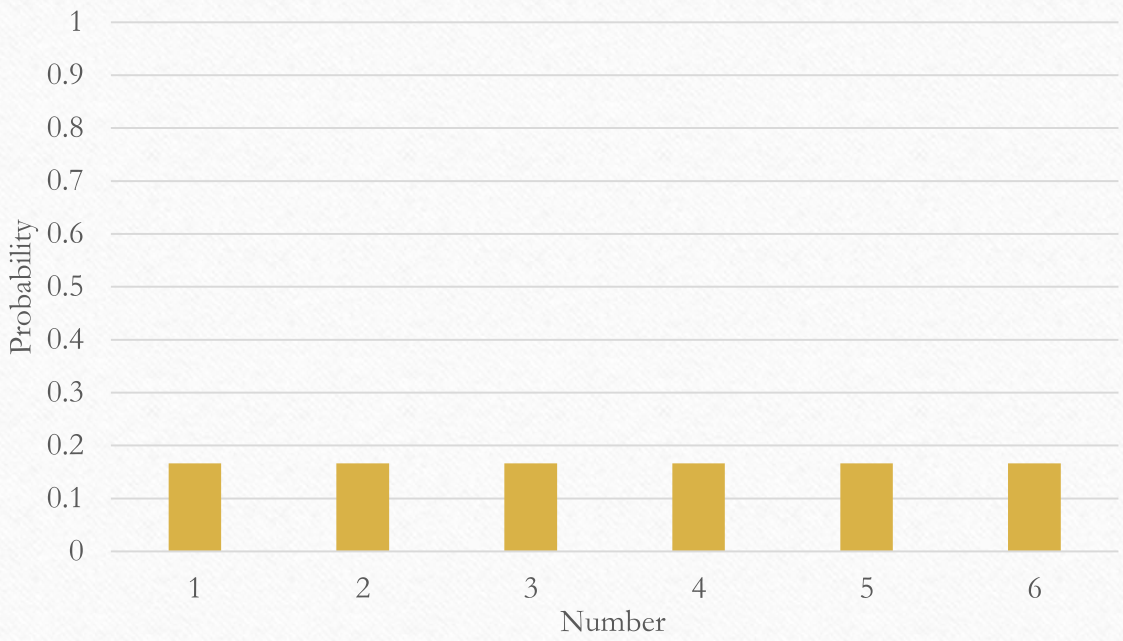
- **Probability distributions** prove useful in scenarios, for instance, when rolling a dice, where the likelihood of every outcome is predetermined.
- Instead of physically rolling a dice 100 times (or more) and constructing a histogram of the results, one can skip the actual rolling.
- By using a **uniform distribution**, we know each potential outcome has a probability of

$$P(X = x) = \frac{1}{6}$$

- Many real-world scenarios around us can be modelled using probability distributions:
  - Number of emails you will receive tomorrow.
  - Exact time your friend will call you tomorrow.
  - Colour of the next car you will buy.
  - Whether it will rain exactly 50 days from now.
  - Number of pages in the next book you read.
  - Number of steps you take in a day.

- A **probability distribution family** serves as a model for a set of scenarios exhibiting similar consistent patterns.
- By utilizing a suitable probability distribution family, we can determine the probabilities for all possible outcomes of a given situation.

Probability Distribution Family



- We've previously come across a type of probability distribution family:
    - The **discrete uniform distribution**!

- Each **probability distribution family** comes with some specific conditions that a scenario should fulfil to fit within that distribution.

- For the **discrete uniform distribution**, two conditions exist:
    - Presence of a discrete random variable.
    - An equal probability for every possible outcome.

- You're sitting on your balcony with a box of chocolates, each with a unique flavour: almond, caramel, mint, orange, dark, white, hazelnut, and raspberry.
- Let X be the flavor of a chocolate you select at random from the box. Can we model X using a discrete uniform distribution?
  - Yes!

- Your preferred chocolate flavour, the hazelnut one, is in the box. You're curious about the probability of picking that specific chocolate.
  - This is the outcome you're keen on, often denoted as k.
- To find $P(X = k)$, we would need to know the total number of chocolates in the box.
  - This is the total number of possible outcomes, often denoted as n.

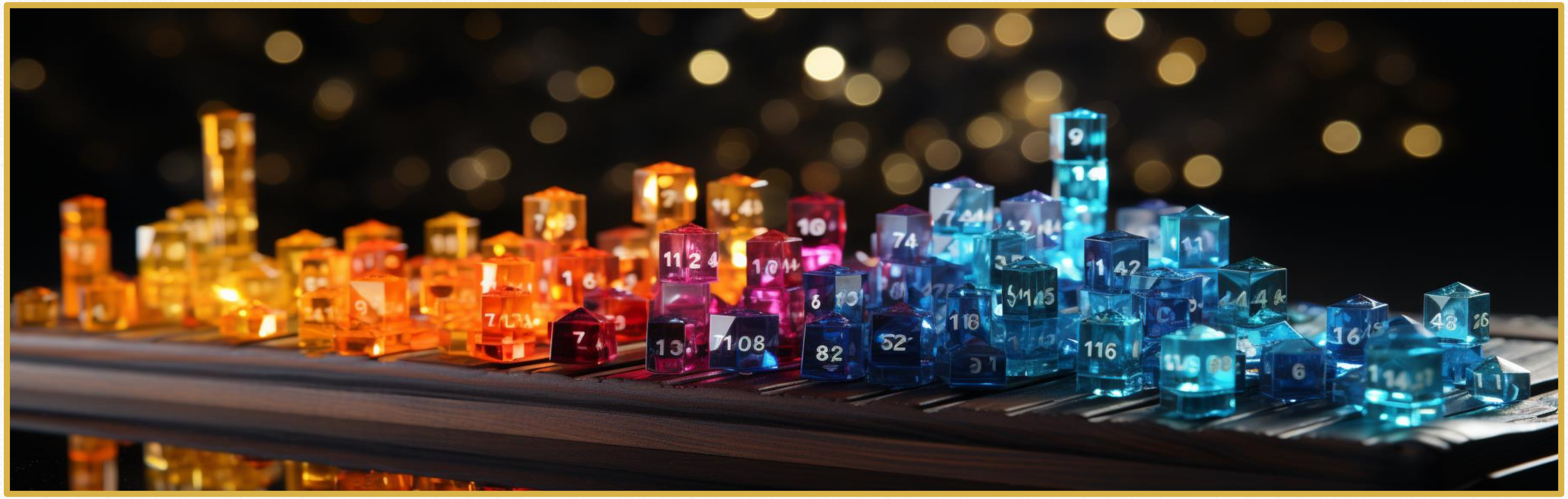
- We refer to n as the family's **parameter**.
- A parameter is a specific value that connects a distribution to a particular scenario. This value guides us in adjusting the distribution to accurately represent the situation.

- Back to our chocolates. We count the total number of chocolates and it has $n = 8$ chocolates.
- Hence $P(X = hazelnut) = \frac{1}{8}$
- From this, we can also learn that each probability distribution family comes with a **probability distribution function**, which helps us in determining specific probability numbers.
- In this case, our probability distribution function is:

$$P(X = k) = \frac{1}{n}$$

Probability Distribution Family

- Every **probability distribution family** will have:
  1. Specific conditions that a scenario must meet for the distribution to be applicable.
  2. Parameters, represented by unique letters for each family.
  3. A dedicated probability distribution function for determining probability values.



Binomial Distributions

- Definition
- Properties of binomial distributions

- The moment is here. An alien invasion has taken over 👽
- You've been training to shoot rifles at these extraterrestrial beings with your buddy Billy. After extensive training, you've determined that you have a 70% chance of hitting your target.

- Every time you scout the area for resources, you always come across 20 aliens. You can only carry 20 bullets for your rifle. If you manage to shoot down 12 aliens, you and Billy can outpace the remaining 8.

- Billy is a statistics expert, so you inquire about the likelihood of successfully shooting 12 out of the 20 aliens during today's mission.

- In this situation, aren’t we looking at a **random variable**?
- Let X represent the number of aliens successfully targeted. To determine the scenario's outcome, we need to evaluate $P(X = 12)$, where X is a **discrete random variable**.

- The situation can be characterized by a **binomial distribution**, which represents the distribution of successes in independent binary (yes/no) experiments conducted with a consistent probability of success.
- Billy explains that every encounter with an alien represents a **trial**. The cumulative number of these trials is denoted by n which serves as the first parameter.
- Successfully targeting an alien is termed a "success", and k represents the specific number of successes under consideration.
- So, for this resources scout, $n = 20$ and $k = 12$

- The second parameter in a binomial distribution is denoted as p, which signifies the likelihood of achieving success in any given trial. In this scenario, success is defined as effectively targeting the alien, with $p = 0.7$.
- When the n trials are independent and has only 2 outcomes, and the success rate for every trial remains consistent at p, the situation can be characterized by a **binomial random variable**.

Parameter Description	**Scenario Value**
Likelihood of achieving success in any given trial	$p = 0.7$
Number of total trials	$n = 20$
Number of successful outcomes of interest	$k = 12$

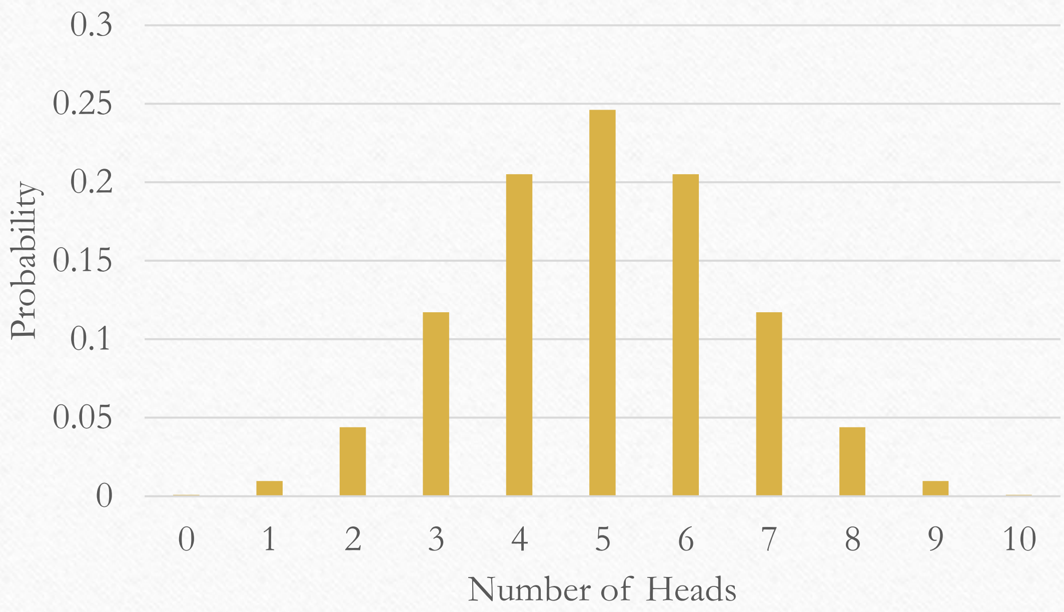
- Billy did some calculations and announces to you that $P(X = 12) = 0.2$ (calculation next slide)
  - Only 20% of hitting 12 aliens 😨
- BUT no worries, because it's okay to hit more than 12 aliens!
  - We should be looking at $P(X \geq 12)$ instead.
  - So we should sum up all the probabilities from $P(X = 12)$ to $P(X = 20)$.

- Formula: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- (Though these days no one really needs such calculations anymore. Just use a computer!)

- Anyway, $P(X \geq 12) = 0.85$
  - You have an 85% chance of hitting 12 or more aliens! 🥴

- Question: Let's say you do a resource scout every day for many months. In the long term, what is the average number of aliens you will hit every resource scout?

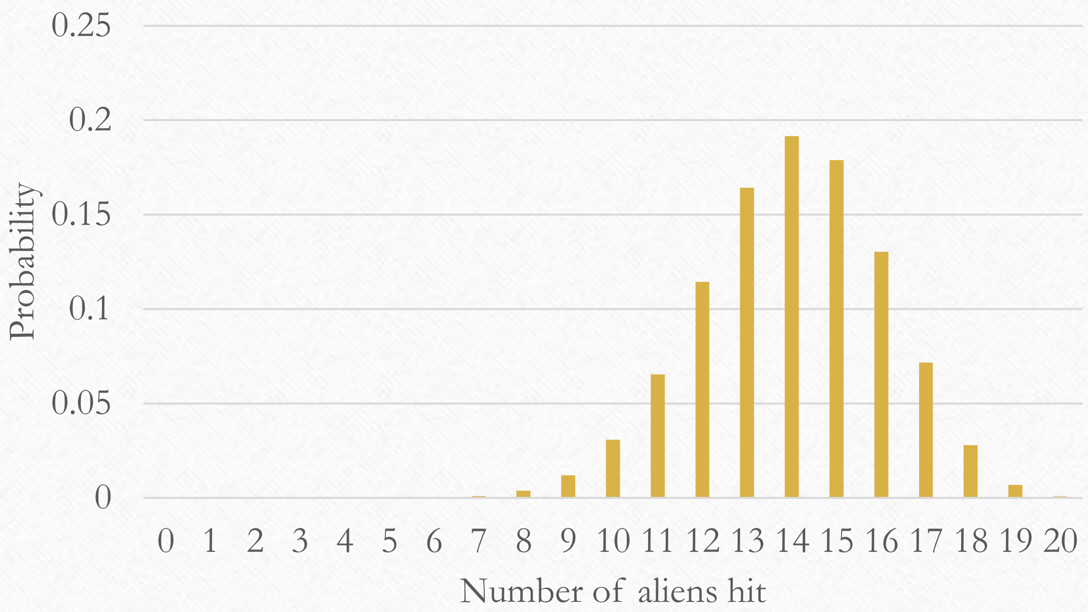
- Here, we are finding the **mean**. The mean of a binomial distribution is $n \times p$.
- So over a long period of time, you will hit $20 \times 0.7 = 14$ aliens per scout.



- Billy illustrates that when plotting a **binomial distribution**, the x-axis represents the possible number of successes, while the y-axis indicates the probability of achieving those successes.
- For instance, referring to the above, if you flip a coin 10 times to try obtaining heads, the x-axis will range from 0 to 10. (Not H or T)

- If you're curious on how the graph is plotted. $n = 10, p = 0.5$

Number of Heads ($\boldsymbol{k}$)	Probability $\boldsymbol{P(X = k)}$
0	0.0009765625
1	0.009765625
2	0.0439453125
3	0.1171875
4	0.205078125
5	0.24609375
6	0.205078125
7	0.1171875
8	0.0439453125
9	0.009765625
10	0.0009765625



- Back to our alien invasion.
- After you collect many months' worth of data and plot it, it displays **binomial distribution** as above.

- Binomial or not?

Scenario	**Binomial or Not?**
Tossing a coin several times and recording whether it lands heads or tails.	Binomial
Drawing a card from a standard deck of 52 cards, recording whether it's a heart, and then replacing it.	Binomial
Drawing a card from a standard deck of 52 cards, recording whether it's a heart, and not replacing it.	Not Binomial
Spinning a wheel divided into 4 equal sections labeled A, B, C, and D, and recording the section where the pointer stops.	Not Binomial
Taking a multiple-choice test where each question has 5 options and guessing the answer to each question.	Binomial

The End