Example: Insurance Deducibles

- Suppose a health insurance company found that 70% of the people they insure stay below their deductible in any given year.
- Each of these people can be thought of as a single trial in a study.
- We label a person a "success" if their healthcare costs do not exceed the deductible.
    - $P(\texttt{success}) = p = 0.7$
    - $P(\texttt{failure}) = 1 - p = 0.3$

The Bernoulli Distribution

- When an individual trial only has two possible outcomes it is called a Bernoulli random variable.
  - These outcomes are often labeled as success or failure.
- *These labels can be completely arbitrary!*
  - We called "not hitting the deductible" a "success", but we could just as well have labeled that the "failure".

The Bernoulli Random Variable

- We code a success as 1 and a failure as 0.
- From here, we can define the mean and standard deviation.

The Bernoulli Random Variable

If X is a random variable that takes the value 1 with probability of success p and 0 with probability $1-p$, then X is a **Bernoulli random variable** with mean

$$\mu = p$$

and standard deviation

$$\sigma = \sqrt{p(1-p)}.$$

The Binomial Distribution

The **binomial distribution** is used to describe the number of successes in a fixed number of trials.

- This is an extension of the Bernoulli distribution.
- We check for a success or failure repeatedly over multiple trials.
- Each *individual* trial can be described with a Bernoulli distribution.

Example: Insurance

- Let's return to the insurance agency where 70% of individuals do not exceed their deductible.
- Suppose the insurance agency is considering a random sample of four individuals they insure.
- What is the probability that exactly one of them will exceed the deductible and the other three will not?

Example

Let's call the four people Ariana (A), Brittany (B), Carlton (C), and Damian (D). Consider a scenario where one person exceeds the deductible:

$$\begin{aligned} & P(A = \texttt{exceed}, B = \texttt{not}, C = \texttt{not}, D = \texttt{not}) \\ & \quad = P(A = \texttt{exceed}) \times P(B = \texttt{not}) \times P(C = \texttt{not}) \times P(D = \texttt{not}) \\ & \quad = (0.3) \times (0.7) \times (0.7) \times (0.7) \\ & \quad = (0.3)^1 \times (0.7)^3 \\ & \quad = 0.103 \end{aligned}$$

Example

- But there are three other scenarios!
    1. Brittany could have been the one to exceed the deductible.
    2. ... or Carlton could have.
    3. ... or Damian.
- In each of these cases, the probability is $(0.7)^3(0.3)^1$.

Example

- These four scenarios consist of all the possible ways that exactly one of these four people could have exceeded the deductible.
- So the total probability is

$$4 \times (0.7)^3 \times (0.3)^1 = 0.412.$$

This is an example of a scenario where we would use a binomial distribution.

Example: Building to Binomial

- There were four people who could have been the single failure.
- Each scenario has the same probability.
- So the final probability was

$$[\#\text{ of scenarios}] \times P(\text{single scenario})$$

Example: Building to Binomial

- The first component of this equation is the number of ways to arrange $k = 3$ successes among $n = 4$ trials.
- The second is the probability of any one of the scenarios.
  - These four scenarios are equally probable.

Building to Binomial

- Consider P(single scenario) with k successes and $n - k$ failures in n trials.
- We know how to handle this!
- We will use the multiplication rule for independent events.

Probability for a Single Scenario

Applying the multiplication rule for independent events,

$$\begin{aligned} P(\text{single scenario}) &= P(k \text{ successes}) \times P(n-k \text{ failures}) \\ &= p \times \cdots \times p \times (1-p) \times \cdots \times (1-p) \\ &= p^k \times (1-p)^{n-k} \end{aligned}$$

This is our general formula for $P(\text{single scenario})$.

Number of Ways to Arrange Successes

The number of ways to arrange k successes and $n-k$ failures is ${}_nC_k$. We also call this the **binomial coefficient**:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

read "n choose k".

Example

We can use this to double check our insurance deductible problem.

Recall that we decided that there were four possible ways to get 3 successes (not exceeding) among 4 people (trials).

$$\begin{aligned}\binom{4}{3} &= \frac{4!}{3!(4-3)!} \\ &= \frac{4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) \times (1)} \\ &= 4\end{aligned}$$

which is just what we decided before!

The Binomial Distribution

Suppose X is a **binomial random variable**. The probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The Binomial Distribution

The expected value (mean) is

$$E(X) = \mu = np$$

and the variance is

$$Var(X) = \sigma^2 = np(1-p)$$

If $p \approx (1-p)$, then the binomial distribution is symmetric.

The Binomial Distribution

We say that X follows a **binomial distribution** with number of trials n and probability of success p if

1. The number of trials is fixed $= n$.
2. The trials are independent.
3. There are two possible outcomes, success/failure.
4. The probability of success is known and fixed $= p$.

Example: Cars at Sac State

Suppose that 38% of Sac State students own a car. A random sample of 20 students is selected. Let X be the number of students in the sample who own a car. What is the distribution of X?

Example: Cars at UCR

Suppose that 38% of Sac State students own a car. A random sample of 20 students is selected. Let X be the number of students in the sample who own a car. What is the distribution of X?

1. $n = 20$ students, so the number of trials is fixed.
2. We have a random sample, so the trials are independent.
3. Success = `car`
   Failure = `no car`
4. $p = P(\texttt{car}) = 0.38$

So $X \sim \text{Bin}(n = 20, p = 0.38)$

Example: Cars at UCR

What is the probability that none of the 20 students own a car?

Example: Cars at UCR

What are the mean and variance of X, the number of students in the sample who own a car?

Computing Binomial Probabilities

1. Check that the (binomial) model is appropriate.
2. Identify n, p, and k.
3. Determine the probability.
4. Interpret the results.

When doing calculations by hand, cancel out as many terms as possible in the binomial coefficient!

Example: Cars at UCR

What is the probability that no more than 2 students own a car?

Example: Cars at UCR

What is the probability that fewer than two students own a car?

Example: Cars at UCR

What is the probability that more than 2 students own a car?