DISTRIBUTIONS OF DISCRETE DATA

ST101 – DR. ARIC LABARR

WHAT ARE DISTRIBUTIONS?

DISTRIBUTIONS OF DISCRETE DATA

RANDOM VARIABLES

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.

RANDOM VARIABLES

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.
  - Finite example: Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}

RANDOM VARIABLES

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.
  - Infinite example: Let x be the number of customers arriving in one day at a small department store where x can take the values of 0, 1, 2, …

RANDOM VARIABLES

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A discrete random variable may assume either a finite number of values or an infinite sequence of values.
- A **continuous random variable** may assume any numerical value in an interval or collection of intervals.

DISCRETE VS. CONTINUOUS

- Discrete Example:
  - Let x be the number of individuals living in a home.

- Continuous Example:
  - Let x be the distance in miles from home to the store.

SUMMARY

- A random variable is a numerical description of the outcome of an experiment.
- A discrete random variable may assume either a finite number of values or an infinite sequence of values.
- A continuous random variable may assume any numerical value in an interval or collection of intervals.

DISCRETE PROBABILITY DISTRIBUTIONS

DISTRIBUTIONS OF DISCRETE DATA

PROBABILITY DISTRIBUTION

- The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable.
- Essentially, what is the frequency of occurrence of different values of the variable.

NOTATION

- **Frequency** – number of observations in each category in the data set
- **Relative Frequency** – proportion of total observations contained in a given category
- **Cumulative Frequency** – summary of data set *i* number of observations with values less than or equal to upper limit of the category
- **Cumulative Relative Frequency** – proportion of observations with value less than or equal to upper limit of the category

PROBABILITY DISTRIBUTION

- The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable.
- Relative frequencies can be used as estimates to the probability of an event occurring.
- Probability distributions for discrete random variables are best described with tables, graphs, or equations.

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90		
1	85		
2	70		
3	45		
4	50		
5	25		
	365		

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	90/365
1	85		
2	70		
3	45		
4	50		
5	25		
	365		

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	0.25
1	85	90+85	0.23
2	70		
3	45		
4	50		
5	25		
	365		

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	0.25
1	85	175	0.23
2	70	245	0.19
3	45	290	0.12
4	50	340	0.14
5	25	365	0.07
	365		**1.00**

DISCRETE PROBABILITY EXAMPLE

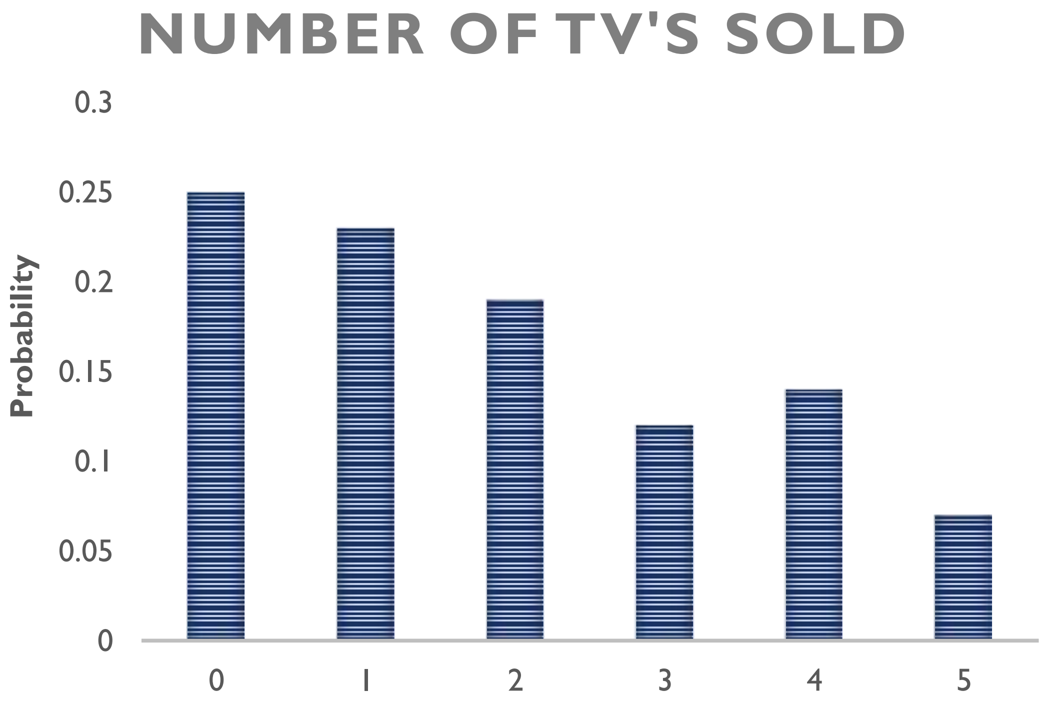
- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	0.25
1	85	175	0.23
2	70	245	0.19
3	45	290	0.12
4	50	340	0.14
5	25	365	0.07
	365		**1.00**

Probability! (→ Relative Frequency)

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.



DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	$P(X = x)$
0	90	90	0.25
1	85	175	0.23
2	70	245	0.19
3	45	290	0.12
4	50	340	0.14
5	25	365	0.07
	365		1.00

SUMMARY

- The probability distribution for a random variable describes how probabilities are distributed over the values of the random variable.
  - Frequency is the number of observations in each category in the data set.
  - Relative frequency is the proportion of total observations contained in a given category.
  - Cumulative frequency is the summary of data set *i* number of observations with values less than or equal to upper limit of the category.
  - Cumulative relative frequency is the proportion of observations with value less than or equal to upper limit of the category.

EXPECTED VALUE AND VARIANCE

DISTRIBUTIONS OF DISCRETE DATA

EXPECTED VALUE

- The **expected value**, or mean, of a random variable is a measure of its central location.
- It is defined by:

$$E(x) = \mu = \sum_{i=1}^{n} x_i \times P(X = x_i)$$

- Think about the expected value as a **weighted mean**, where the probability function serves as the weight.

EXPECTED VALUE

- The **expected value**, or mean, of a random variable is a measure of its central location.
- It is defined by:

$$E(x) = \mu = \sum_{i=1}^{n} x_i \times P(X = x_i)$$

- Think about the expected value as a **weighted mean**, where the probability function serves as the weight.
- What if the everything had the same probability of $1/n$?

EXPECTED VALUE

- The **expected value**, or mean, of a random variable is a measure of its central location.
- It is defined by:

$$E(x) = \mu = \sum_{i=1}^{n} x_i \times \frac{1}{n}$$

- Think about the expected value as a **weighted mean**, where the probability function serves as the weight.
- What if the everything had the same probability of $1/n$?

EXPECTED VALUE

- The **expected value**, or mean, of a random variable is a measure of its central location.
- It is defined by:

$$E(x) = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Think about the expected value as a **weighted mean**, where the probability function serves as the weight.
- What if the everything had the same probability of $1/n$?

EXPECTED VALUE

- The **expected value**, or mean, of a random variable is a measure of its central location.
- It is defined by:

$$E(x) = \mu = \sum_{i=1}^{n} x_i \times P(X = x_i)$$

- Think about the expected value as a weighted mean, where the probability function serves as the weight.

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	$P(X = x)$	$x_i \times P(X = x_i)$
0	90	90	0.25	
1	85	175	0.23	
2	70	245	0.19	
3	45	290	0.12	
4	50	340	0.14	
5	25	365	0.07	
	365		**1.00**	

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	$P(X = x)$	$x_i \times P(X = x_i)$
0	90	90	0.25	0.00
1	85	175	0.23	
2	70	245	0.19	
3	45	290	0.12	
4	50	340	0.14	
5	25	365	0.07	
	365		**1.00**	

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	$P(X = x)$	$x_i \times P(X = x_i)$
0	90	90	0.25	0.00
1	85	175	0.23	0.23
2	70	245	0.19	
3	45	290	0.12	
4	50	340	0.14	
5	25	365	0.07	
	365		**1.00**	

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	$P(X = x)$	$x_i \times P(X = x_i)$
0	90	90	0.25	0.00
1	85	175	0.23	0.23
2	70	245	0.19	0.38
3	45	290	0.12	0.36
4	50	340	0.14	0.56
5	25	365	0.07	0.35
	365		**1.00**	**1.88**

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- We expect to sell 1.88 TV's per day on average.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	$P(X = x)$	$x_i \times P(X = x_i)$
0	90	90	0.25	0.00
1	85	175	0.23	0.23
2	70	245	0.19	0.38
3	45	290	0.12	0.36
4	50	340	0.14	0.56
5	25	365	0.07	0.35
	365		**1.00**	**1.88**

VARIANCE

- The **variance** of a random variable is a measure of its variability/spread.
- It is defined by:

$$Var(x) = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i)$$

- The **standard deviation** is the square root of the variance.

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- We expect to sell 1.88 TV's per day on average.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88		
1	85	0.23	0.23			
2	70	0.19	0.38			
3	45	0.12	0.36			
4	50	0.14	0.56			
5	25	0.07	0.35			
	365	**1.00**	**1.88**			

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- We expect to sell 1.88 TV's per day on average.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88		
1	85	0.23	0.23	-0.88		
2	70	0.19	0.38	0.12		
3	45	0.12	0.36	1.12		
4	50	0.14	0.56	2.12		
5	25	0.07	0.35	3.12		
	365	**1.00**	**1.88**			

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- We expect to sell 1.88 TV's per day on average.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88	3.53	
1	85	0.23	0.23	-0.88		
2	70	0.19	0.38	0.12		
3	45	0.12	0.36	1.12		
4	50	0.14	0.56	2.12		
5	25	0.07	0.35	3.12		
	365	**1.00**	**1.88**			

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- We expect to sell 1.88 TV's per day on average.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88	3.53	
1	85	0.23	0.23	-0.88	0.77	
2	70	0.19	0.38	0.12	0.01	
3	45	0.12	0.36	1.12	1.25	
4	50	0.14	0.56	2.12	4.49	
5	25	0.07	0.35	3.12	9.73	
	365	**1.00**	**1.88**			

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- We expect to sell 1.88 TV's per day on average.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88	3.53	0.883
1	85	0.23	0.23	-0.88	0.77	
2	70	0.19	0.38	0.12	0.01	
3	45	0.12	0.36	1.12	1.25	
4	50	0.14	0.56	2.12	4.49	
5	25	0.07	0.35	3.12	9.73	
	365	**1.00**	**1.88**			

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- The variance of daily sales is 2.522 TV's squared.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88	3.53	0.883
1	85	0.23	0.23	-0.88	0.77	0.177
2	70	0.19	0.38	0.12	0.01	0.002
3	45	0.12	0.36	1.12	1.25	0.150
4	50	0.14	0.56	2.12	4.49	0.629
5	25	0.07	0.35	3.12	9.73	0.681
	365	**1.00**	**1.88**			**2.522**

DISCRETE PROBABILITY EXAMPLE

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- The standard deviation of daily sales is 1.588 TV's.

TV's Sold	Number of Days (Freq)	$P(X = x)$	$x_i \times P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
0	90	0.25	0.00	-1.88	3.53	0.883
1	85	0.23	0.23	-0.88	0.77	0.177
2	70	0.19	0.38	0.12	0.01	0.002
3	45	0.12	0.36	1.12	1.25	0.150
4	50	0.14	0.56	2.12	4.49	0.629
5	25	0.07	0.35	3.12	9.73	0.681
	365	**1.00**	**1.88**			**2.522**

SUMMARY

- The expected value, or mean, of a random variable is a measure of its central location:

$$E(x) = \mu = \sum_{i=1}^{n} x_i \times P(X = x_i)$$

- The variance of a random variable is a measure of its variability/spread:

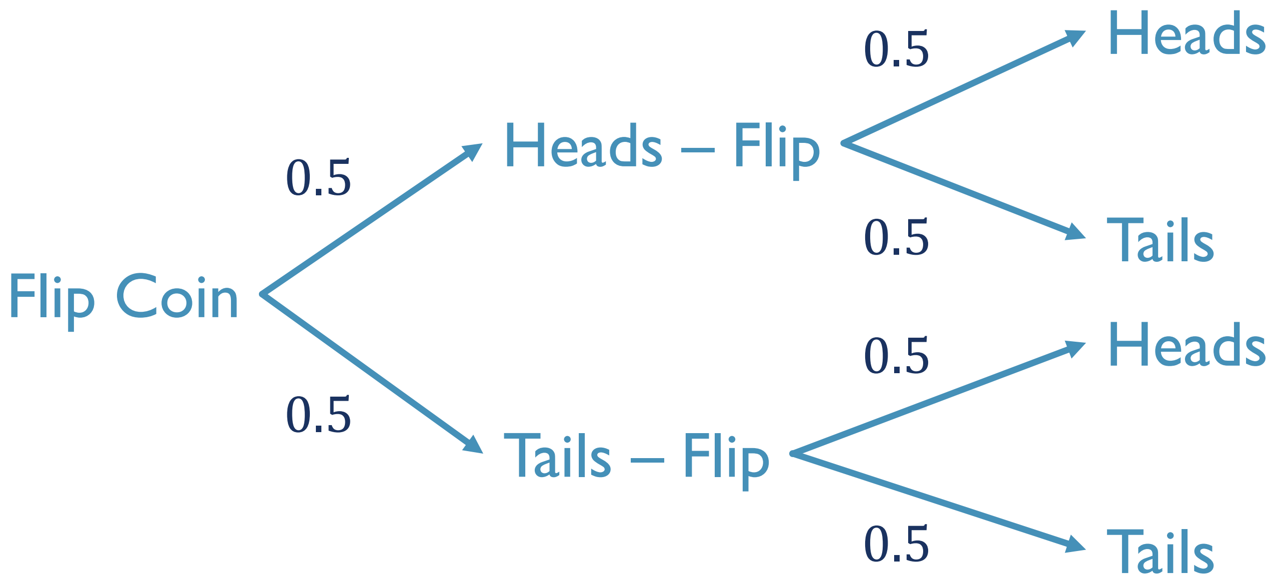
$$Var(x) = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i)$$

BINOMIAL DISTRIBUTION

DISTRIBUTIONS OF DISCRETE DATA

EXAMPLE REVIEW

- Example: you have a 2-step random process where you flip a coin twice (independent flips):



BINOMIAL EXPERIMENT

- There are 4 properties of a binomial experiment:

1. The experiment consists of a sequence of n identical trials.
2. Only two outcomes, success or failure, are possible on each trial.
3. The probability of a success, denoted as p, does not change from trial to trial.
4. The trials are independent.

BINOMIAL EXPERIMENT

- There are 4 properties of a binomial experiment:
  1. The experiment consists of a sequence of *n* identical trials. (2 coin flips)
  2. Only two outcomes, success or failure, are possible on each trial. (H or T)
  3. The probability of a success, denoted as *p*, does not change from trial to trial. (0.5)
  4. The trials are independent. (Independent coin flips)

BINOMIAL DISTRIBUTION

- The **binomial distribution** looks at the probabilities of the number of successes occurring in the *n* trials.
- We use x to denote the number of successes occurring in the *n* trials.

BINOMIAL DISTRIBUTION

- The **binomial distribution** looks at the probabilities of the number of successes occurring in the *n* trials.
- We use x to denote the number of successes occurring in the *n* trials.
- For example:
  - Let a success be rolling a dice and getting a 2.
  - We roll a dice 10 times, *n*, and successfully role a 2, three times, x.
  - We are interested in the probability of exactly 3 rolls equal to 2.

PROBABILITY FUNCTION

- The **binomial probability function** is defined as:

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Number of outcomes providing exactly x successes in *n* trials

PROBABILITY FUNCTION

- The **binomial probability function** is defined as:

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Number of outcomes providing exactly x successes in *n* trials

Probability of a particular sequence of trial outcomes with x successes in *n* trials

BINOMIAL DISTRIBUTION EXAMPLE

- For example:
  - Let a success be rolling a dice and getting a 2.
  - We roll a dice 10 times, *n*, and successfully role a 2, three times, x.
  - We are interested in the probability of exactly 3 rolls equal to 2.

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

BINOMIAL DISTRIBUTION EXAMPLE

- For example:
  - Let a success be rolling a dice and getting a 2.
  - We roll a dice 10 times, *n*, and successfully role a 2, three times, x.
  - We are interested in the probability of exactly 3 rolls equal to 2.

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

$$f(x=3) = \frac{10!}{3!\,(10-3)!} \left(\frac{1}{6}\right)^3 \left(1-\frac{1}{6}\right)^{10-3}$$

BINOMIAL DISTRIBUTION EXAMPLE

- For example:
  - Let a success be rolling a dice and getting a 2.
  - We roll a dice 10 times, *n*, and successfully role a 2, three times, x.
  - We are interested in the probability of exactly 3 rolls equal to 2.

$$f(x) = \frac{n!}{x!\,(n-x)!}p^x(1-p)^{n-x}$$

$$f(x=3) = \frac{10!}{3!\,(10-3)!}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^{10-3}$$

$$f(x=3) = 120 \times 0.0013 = 0.155$$

BINOMIAL DISTRIBUTION EXAMPLE

- For example:
  - Let a success be rolling a dice and getting a 2.
  - We roll a dice 10 times, *n*, and successfully role a 2, three times, x.
  - We are interested in the probability of exactly 3 rolls equal to 2.

$$f(x) = \frac{n!}{x!\,(n-x)!}p^x(1-p)^{n-x}$$

$$f(x=3) = \frac{10!}{3!\,(10-3)!}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^{10-3}$$

$$f(x=3) = 120 \times 0.0013 = 0.155$$

BINOMIAL DISTRIBUTION EXAMPLE

- You have a retention rate of 90% for your employees annually.
- In other words, any random employee has a probability of 0.1 to leave this year.
- Choosing 3 employees at random, what is the probability that exactly 1 of them will leave the company this year?

BINOMIAL DISTRIBUTION EXAMPLE

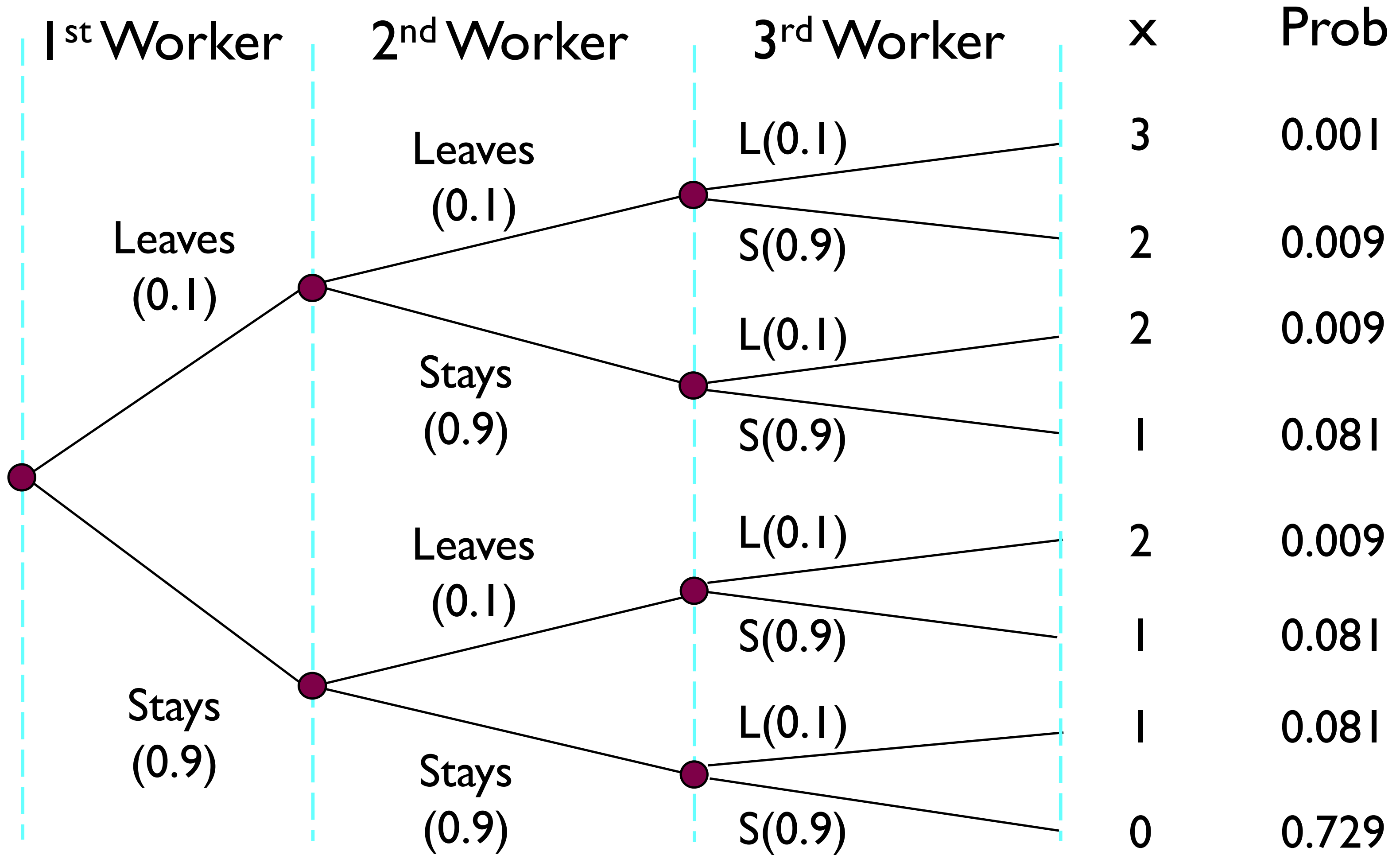
- You have a retention rate of 90% for your employees annually.
- In other words, any random employee has a probability of 0.1 to leave this year.
- Choosing 3 employees at random, what is the probability that exactly 1 of them will leave the company this year?

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

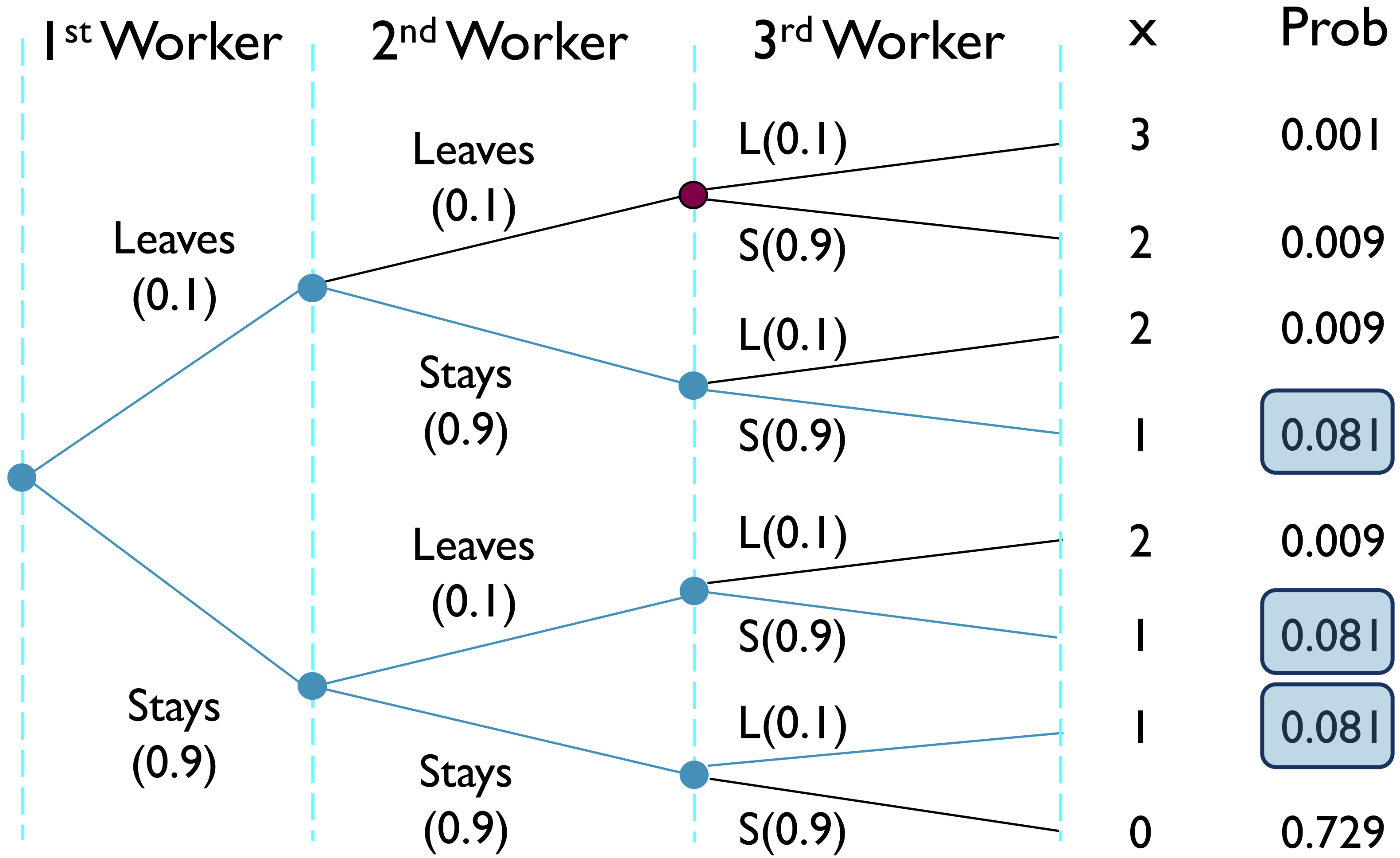
$$f(x=1) = \frac{3!}{1!\,(3-1)!} (0.1)^1 (1-0.1)^{3-1}$$

$$f(x=1) = 3 \times 0.081 = 0.243$$

BINOMIAL DISTRIBUTION EXAMPLE



BINOMIAL DISTRIBUTION EXAMPLE



EXPECTED VALUE AND VARIANCE/STANDARD DEVIATION

- For the binomial distribution, the following is always true:
- Expected value:

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

- Variance:

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

- Standard deviation:

$$SD(x) = \sigma = \sqrt{np(1-p)}$$

BINOMIAL DISTRIBUTION EXAMPLE

- For example:
  - Let a success be rolling a dice and getting a 2.
  - We roll a dice 10 times, *n*, and successfully role a 2, three times, x.
  - We are interested in the probability of exactly 3 rolls equal to 2.

$$E(x) = \mu = np = 10 \times \left(\frac{1}{6}\right) = 1.67$$

We **expect** to roll a 2 on the dice 1.67 times out of 10 chances.

BINOMIAL DISTRIBUTION EXAMPLE

- For example:
- You have a retention rate of 90% for your employees annually.
- In other words, any random employee has a probability of 0.1 to leave this year.
- Choosing 3 employees at random, what is the probability that exactly 1 of them will leave the company this year?

$$E(x) = \mu = np = 3 \times (0.1) = 0.3$$

We **expect** to roll a 0.3 of the 3 employees to leave this year.

SUMMARY

- The binomial distribution looks at the probabilities of the number of successes occurring in the *n* independent trials.
- The binomial probability function is comprised of two intuitive pieces:

$$f(x) = \frac{n!}{x!\,(n-x)!}\,p^{x}(1-p)^{n-x}$$

Number of outcomes providing exactly x successes in *n* trials

Probability of a particular sequence of trial outcomes with x successes in *n* trials