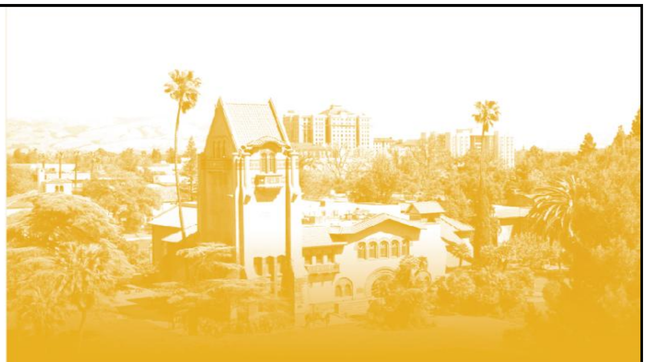


SJSU SAN JOSÉ STATE UNIVERSITY

Random Variables
& Discrete Distributions

1

SJSU SAN JOSÉ STATE UNIVERSITY

Agenda

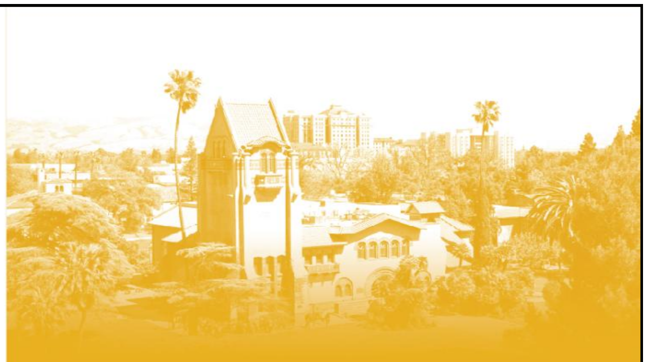
- Random Variables
- Discrete Random Variables & Probability Distributions
  - Bernoulli Distribution
  - Binomial Distribution
  - Poisson Distribution
- Cumulative Distribution Functions

2

2

SJSU SAN JOSÉ STATE UNIVERSITY

Random Variables



3

SJSU SAN JOSÉ STATE UNIVERSITY

What is a Random Variable?

A random variable is a numeric quantity whose value depends on the outcome of a <u>random event</u>.

- We normally use a capital letter, like X, to denote a random variable
- The values of a random variable are denoted with a lowercase letter, in this case x
  - e.g. P(X = x) – denotes the probability of the (random variable) X having the value x.

4

4

SJSU SAN JOSÉ STATE UNIVERSITY

Types of Random Variables

There are two types of random variables:

- Discrete random variables often take only integer values
- Continuous random variables take real (decimal) values

Feature	Continuous Random Variable	Discrete Random Variable
Possible Values	Infinite (uncountable)	Countable (finite or infinite)
Probability at a Single Value	0	Can be nonzero
Described by	Probability Density Function (PDF)	Probability Mass Function (PMF)
Example	Height, time, weight	Dice rolls, # of students in a class

SJSU SAN JOSÉ STATE UNIVERSITY

Discrete Random Variables

A discrete random variable is a type of random variable that can take on a finite or countably infinite set of distinct values.

- Countable values
- These values are typically whole numbers and are obtained from a random experiment.

Examples:

- Number of heads in 3 coin flips → {0, 1, 2, 3}
- Number of students in a classroom → {0, 1, 2, ...}
- Number of cars passing a toll booth in an hour → {0, 1, 2, ...}
- Roll of a die → {1, 2, 3, 4, 5, 6}

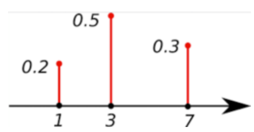
SJSU SAN JOSÉ STATE UNIVERSITY

Probability Distributions of Discrete Random Variables

A probability distribution of a discrete random variable is a list of its possible values and the probabilities that it takes on those values.

Value of X	x_1	x_2	x_3	...
Probability	p_1	p_2	p_3	...

$$P(X = x_i) = p_i$$



- It's also known as the probability mass function (PMF).
- This must satisfy two conditions:

$$0 \le p_i \le 1 \; for \; all \; i$$

$$p_1 + p_2 + \cdots = 1$$

SJSU SAN JOSÉ STATE UNIVERSITY

Example

Let X be the # of heads obtained in 2 tosses of a fair coin. What is the distribution of X?

Value of X	0	1	2
Outcomes	TT	HT, TH	HH
Probability	1/4	1/2	1/4

SJSU SAN JOSÉ STATE UNIVERSITY

Example: A Card Game

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. What is the probability distribution of your earning?

Outcome	X	P(X)

9

9

SJSU SAN JOSÉ STATE UNIVERSITY

Expected Value of a Discrete Random Variable

- We are often interested in the average outcome of a random variable.
- We call this the expected value (mean), and it is a weighted average of the possible outcomes

$$\mu_x = E(X) = \sum_{i=1}^{k} x_i \, P(X = x_i)$$

10

10

SJSU SAN JOSÉ STATE UNIVERSITY

Example: A Card Game Revisited

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. What is your expected earning?

Outcome	X	$P(X)$	$X \cdot P(X)$
Heart (not ace)	1	12/52	
Ace	5	4/52	
King of spades	10	1/52	
All else	0	35/42	

SJSU SAN JOSÉ STATE UNIVERSITY

Interpretation of Expected Values

- If we play the card game a huge number of time, what's the average earning per game?

- If it charges a certain amount of money each time to play the game, what is the maximum amount you would be willing to pay? Explain your reasoning.

- What is a fair game?

SJSU SAN JOSÉ STATE UNIVERSITY

Variance of a Random Variable

- We are also often interested in the variability in the values of a random variable.
- The variance of a random variable X, denoted as ${\sigma_X}^2$, is defined as:

$$ {\sigma_X}^2 = Var(X) = \sum_{i=1}^{k} (x_i - E(X))^2 \, P(X = x_i) $$

- The standard deviation of a random variable X, denoted as σ_X, is simply the square root of the variance.

SJSU SAN JOSÉ STATE UNIVERSITY

Example: A Card Game Revisited

For the previous card game example, how much would you expect the winnings to vary from game to game?

Outcome	X	$P(X)$	$X \cdot P(X)$	$(X - E(X))^2$	$P(X) \cdot (X - E(X))^2$
Heart (not ace)	1	12/52	12/52		
Ace	5	4/52	20/52		
King of spades	10	1/52	10/52		
All else	0	35/42	0		

SJSU SAN JOSÉ STATE UNIVERSITY

Linear Combinations of Random Variables

Let X, Y, Z be random variables and a, b, c be any fixed numbers.

- Expected Value of a linear combination of random variables:

$$E(aX + bY + cZ + \cdots) = a \cdot E(X) + b \cdot E(Y) + c \cdot E(Z) + \cdots$$

Note: This is always valid

- Variance of a linear combination of random variables:

$$Var(aX + bY + cZ + \cdots) = a^2 \cdot Var(X) + b^2 \cdot Var(Y) + c^2 \cdot Var(Z) + \cdots$$

Note: If the random variables are not independent, the variance calculation gets a little more complicated and is beyond the scope of this course.

15

15

SJSU SAN JOSÉ STATE UNIVERSITY

Example

On average you take 10 minutes for each statistics homework problem and 15 minutes for each chemistry homework problem. This week you have 5 statistics and 4 chemistry homework problems assigned.

What is the total time you expect to spend on statistics and chemistry homework for the week?

16

16

SJSU SAN JOSÉ STATE UNIVERSITY

Example

The standard deviation of the time you take for each statistics homework problem is 1.5 minutes, and it is 2 minutes for each chemistry problem.

What is the standard deviation of the time you expect to spend on statistics and chemistry homework for the week if you have 5 statistics and 4 chemistry homework problems assigned?

17

17

SJSU SAN JOSÉ STATE UNIVERSITY

Simplifying Random Variables

Random variables do not work like normal algebraic variables:

$$X + X \neq 2X$$

$$E(X + X) = E(X) + E(X) = 2\,E(X)$$

$$Var(X + X) = Var(X) + Var(X) = 2\,Var(X)$$

$$E(2X) = 2\,E(X)$$

$$Var(2X) = 2^2\,Var(X) = 4\,Var(X)$$

$$\boxed{E(X + X) = E(2X) \quad \textbf{but} \quad Var(X + X) \neq Var(2X)}$$

18

18

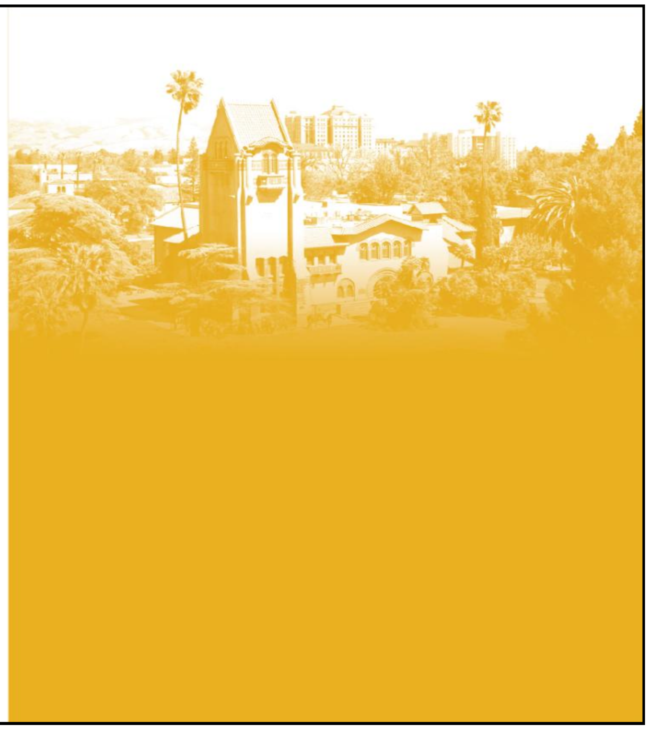
Practice

A casino game costs $5 to play. If you draw first a red card, then you get to draw a second card. If the second card is the ace of hearts, you win $500. If not, you don't win anything, i.e. lose your $5. What is your expected profits (or losses) from playing this game? Note: profit (or loss) = winnings - cost.

19

19

Distributions for Discrete Random Variables



20

SJSU SAN JOSÉ STATE UNIVERSITY

The Bernoulli Distribution

The Bernoulli distribution is a discrete probability distribution for a binary (two-outcome) experiment.

- This is the simplest discrete probability distribution that's based on a single parameter p
- It models a single trial with only two possible outcomes:
  - Success (1) with probability p
  - Failure (0) with probability $1-p$
- Example:

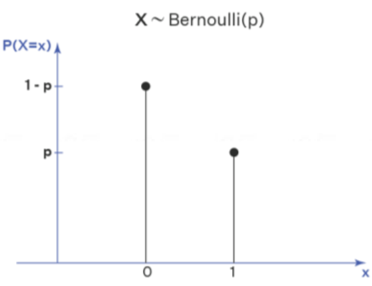
  Tossing a coin (Heads = 1, Tails = 0)

SJSU SAN JOSÉ STATE UNIVERSITY

Probability Mass Function of the Bernoulli Distribution

The probability mass function (PMF) of a Bernoulli random variable X is:

$$P(X=x) = \underbrace{p^x}_{\text{probability of success } p}\ \underbrace{(1-p)^{1-x}}_{\text{probability of failure } 1-p} \qquad x \in \{0,1\}$$



Example: If $p = 0.7$, then:

- $P(X=1) = 0.7$
- $P(X=0) = 1 - 0.7 = 0.3$

SJSU SAN JOSÉ STATE UNIVERSITY

Mean and Variance of the Bernoulli Distribution

- Mean (Expected Value):

$$E(X) = p$$

- Variance

$$Var(X) = p\,(1 - p)$$

23

23

SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Bernoulli Distribution

- Coin flips: Determining the likelihood of heads or tails.
- Quality control: Checking if a manufactured product is defective (yes/no).
- Medical testing: Outcome of a drug being effective or not.
- Marketing: Whether a customer clicks on an ad (click/no-click).

24

24

The Binomial Distribution

The Binomial distribution models the number of successes in n independent Bernoulli trials.

- It's defined by two parameters:
  - # of trials n
  - probability of success p

Example:

- Tossing a coin n times and counting the number of heads.

Probability Mass Function of the Binomial Distribution

The probability mass function (PMF) of a Binomial random variable X is:

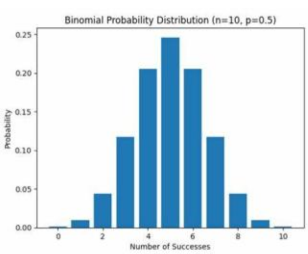
$$P(X=k) = \binom{n}{k} p^k \,(1-p)^{n-k}$$

probability of observing $\boldsymbol{k}$ occurrences

probability of k successes $\boldsymbol{p}$

probability of $n-k$ failures $\boldsymbol{1-p}$

Example: Rolling a fair die 5 times and how many times we get a 6.



SJSU SAN JOSÉ STATE UNIVERSITY

Mean and Variance of the Binomial Distribution

- Mean (Expected Value):

$$E(X) = n\,p$$ expected number of successes in **n** trials.

- Variance

$$Var(X) = n\,p\,(1-p)$$ the spread of the number of successes

27

27

SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Binomial Distribution

- Manufacturing: Counting defective items in a batch.
- Elections: Predicting how many people vote for a candidate in a sample.
- Sports: Number of successful shots made in a basketball game.
- Medical trials: Patients responding positively to a new drug.

28

28

SJSU SAN JOSÉ STATE UNIVERSITY

The Poisson Distribution

The Poisson distribution models the number of events occurring in a fixed interval of time or space, given a constant average rate of occurrence.

- Used when events happen independently and randomly over time.
- It's defined by one parameter:
  - Expected (mean) # of occurrences λ

Examples:

- # of customer arrivals at a store per hour.
- # of emails received per day.
- # of defects in a meter of fabric.

SJSU SAN JOSÉ STATE UNIVERSITY

Probability Mass Function of the Poisson Distribution

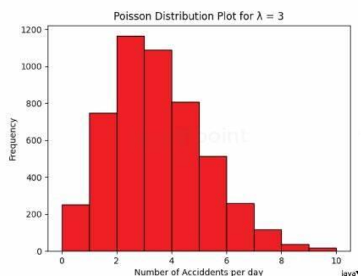
The probability mass function (PMF) of a Poisson random variable X is:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

probability of observing $\boldsymbol{k}$ occurrences over a given interval

$\boldsymbol{\lambda}$ - expected/average # of events

$\boldsymbol{k}$ – desired # of events



SJSU SAN JOSÉ STATE UNIVERSITY

Mean and Variance of the Poisson Distribution

- Mean (Expected Value):

$$E(X) = \lambda$$ expected number of events in the interval

- Variance

$$Var(X) = \lambda$$ the spread of the number of successes

31

31

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Call Center

A call center receives an average of 3 calls every 10 minutes. Assuming the number of calls follows a Poisson distribution, find the probability that the call center will receive exactly 5 calls in the next 10 minutes.

32

32

SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Poisson Distribution

- Traffic flow: Number of cars passing a toll booth per minute.
- Healthcare: Number of patients arriving at an emergency room per hour.
- Telecommunications: Number of dropped calls in a network per day.
- Manufacturing: Number of defects found per batch of products.

33

33

SJSU SAN JOSÉ STATE UNIVERSITY

The Uniform Distribution

The Uniform distribution is a discrete probability distribution where a finite number of values n are equally likely to be observed.

- Each outcome has the same probability: $\frac{1}{n}$
- Example:

  Rolling a fair six-sided dice (1 to 6) ➔ each outcome has a probability of 1/6.

34

34

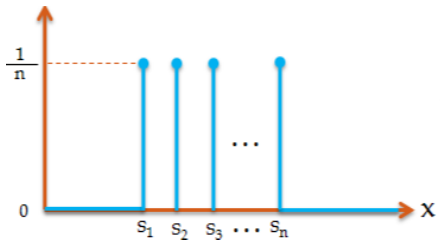
SJSU SAN JOSÉ STATE UNIVERSITY

Probability Mass Function of the Uniform Distribution

The probability mass function (PMF) of a Uniform random variable X is:

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

n # of possible outcomes



SJSU SAN JOSÉ STATE UNIVERSITY

Mean and Variance of the Uniform Distribution

- Mean (Expected Value):

$$E(X) = \frac{x_1 + x_n}{2}$$

- Variance

$$Var(X) = \frac{n^2 - 1}{12}$$

SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Uniform Distribution

- Random Sampling: Used in simulations and Monte Carlo methods.
- Gaming & Gambling: Dice rolls, lotteries, and card games.
- Cryptography: Key generation and security protocols.
- Fair Decision-Making: Randomized selection processes, such as lotteries or raffle draws.

37

37

SJSU SAN JOSÉ STATE UNIVERSITY

Summary of the 3 Discrete Distributions

- Bernoulli Distribution: Models a single trial with two outcomes.
  - Coin flips, success/failure experiments
- Binomial Distribution: Models the number of successes in independent Bernoulli trials.
  - Number of defective products in a batch, number of heads in coin flips
- Poisson Distribution: Models the number of events in a fixed interval, given a known average rate.
  - # of calls to a call center per hour, # of customer arrivals per minute

38

38

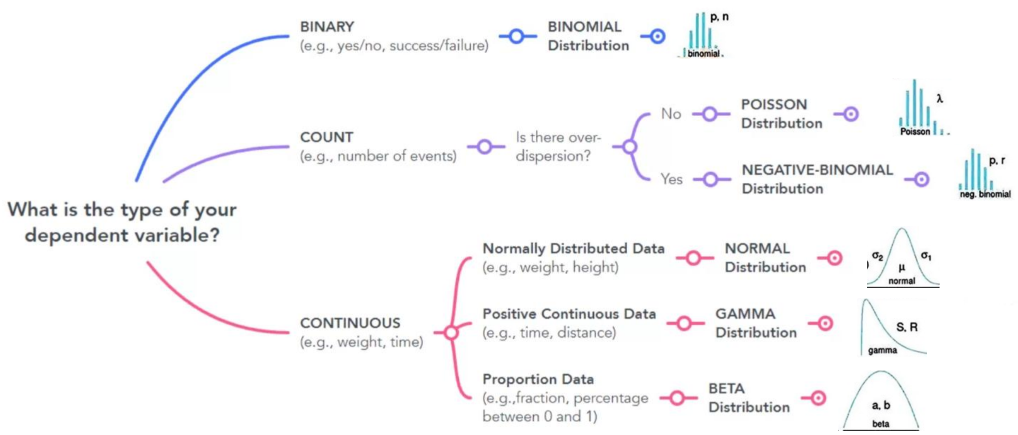
SJSU SAN JOSÉ STATE UNIVERSITY

Summary Table of the 4 Discrete Distributions

Distribution	Definition	Parameters	PMF	Mean	Variance
Bernoulli	Models a single binary outcome	p	$P(X=x) = p^x\,(1-p)^{1-x}$	$E(X) = p$	$Var(X) = p\,(1-p)$
Binomial	Models the # of successes in independent Bernoulli trials	n, p	$P(X=k) = \binom{n}{k} p^k\,(1-p)^{n-k}$	$E(X) = n\,p$	$Var(X) = n\,p\,(1-p)$
Poisson	Models the # of events in a fixed interval	λ	$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$	$E(X) = \lambda$	$Var(X) = \lambda$
Uniform	Models n equally likely outcomes	n	$P(X=k) = \frac{1}{n}$	$E(X) = \frac{x_1 + x_n}{2}$	$Var(X) = \frac{n^2-1}{12}$

SJSU SAN JOSÉ STATE UNIVERSITY

Random Variables & Distributions



https://statisticseasily.com/generalized-linear-model-distribution-and-link-function/

SAN JOSÉ STATE UNIVERSITY

Cumulative Distributions for Discrete Random Variables

41

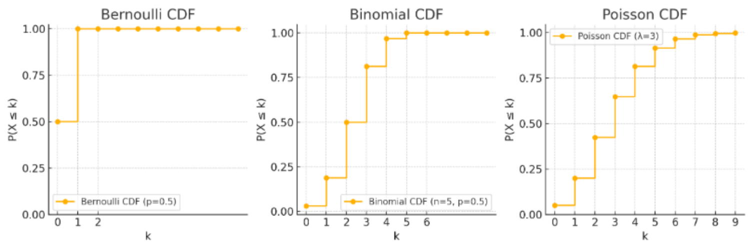
SAN JOSÉ STATE UNIVERSITY

Cumulative Distribution Function (CDF) for Discrete Distributions

The Cumulative Distribution Function (CDF) of a discrete random variable X gives the probability that X takes a value less than or equal to a given number k. It is defined as:

$$F(k) = P(X \leq k) = \sum_{i=-\infty}^{k} P(X = i)$$

probability mass function (PMF)



42

42

SJSU SAN JOSÉ STATE UNIVERSITY

Cumulative Distribution Function of the Bernoulli Distribution

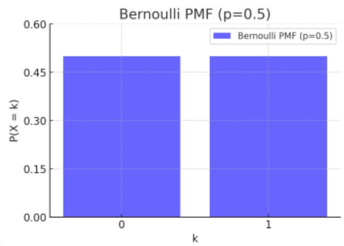
The Cumulative Distribution Function (CDF) of a Bernoulli random variable X is:

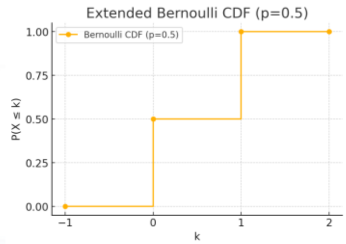
$$F(k) = \begin{cases} 0, & k < 0 \\ 1-p & 0 \le k \le 1 \\ 1, & k > 1 \end{cases}$$

probability of success $\boldsymbol{p}$

- The CDF is a step function that jumps at $X = 0$ and $X = 1$.
- Example: For $p = 0.5$:

k	0	1
$P(X = k)$	0.5	0.5
$F(k)$	0.5	1.0





43

SJSU SAN JOSÉ STATE UNIVERSITY

Cumulative Distribution Function of the Binomial Distribution

The Cumulative Distribution Function (CDF) of a Binomial random variable X is:

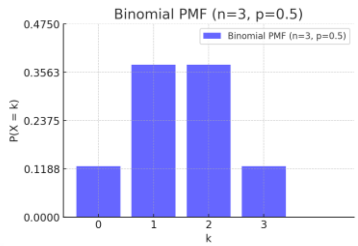
$$F(k) = P(X \le k) = \sum_{i=0}^{k} \binom{n}{i} p^i \, (1-p)^{n-i}$$

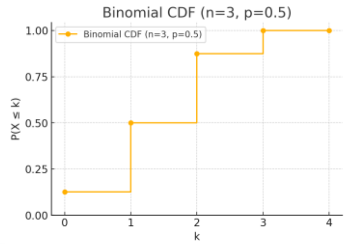
probability of success $\boldsymbol{p}$

- There is no closed-form formula for this summation, so this is usually computed numerically.

Example: For $\mathrm{n} = 3, p = 0.5$:

k	0	1	2	3
$P(X = k)$	0.125	0.375	0.375	0.125
$F(k)$	0.125	0.500	0.875	1.0





44

SJSU SAN JOSÉ STATE UNIVERSITY

Cumulative Distribution Function of the Poisson Distribution

The Cumulative Distribution Function (CDF) of a Poisson random variable X is:

$$F(k) = P(X \le k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^{i}}{i!}$$

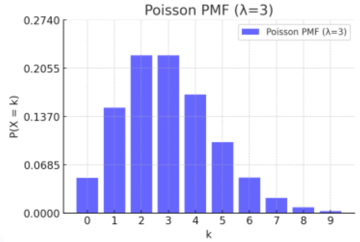
λ - expected/average # of events
k – desired # of events

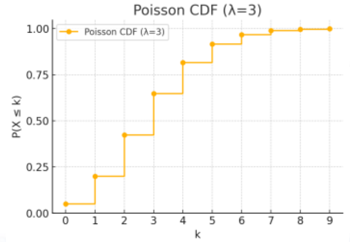
- There is no closed-form formula for this summation, so this is usually computed numerically.

Example: For $\lambda = 3, p = 0.5$:

k	0	1	2	3	4
$P(X = k)$	0.0498	0.1494	0.2240	0.1680	0.1680
$F(k)$	0.0498	0.1992	0.4232	0.6472	0.8152

k	5	6	7	8	9
$P(X = k)$	0.1008	0.0504	0.0216	0.0081	0.0027
$F(k)$	0.9161	0.9665	0.9881	0.9962	0.9989





SJSU SAN JOSÉ STATE UNIVERSITY

Cumulative Distribution Function of the Uniform Distribution

The Cumulative Distribution Function (CDF) of a Uniform random variable X is:

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

Example: For $\mathrm{n} = 6$: