IIT ROORKEE

FREE ONLINE EDUCATION swayam शिक्षित भारत, उन्नत भारत

NPTEL ONLINE CERTIFICATION COURSE

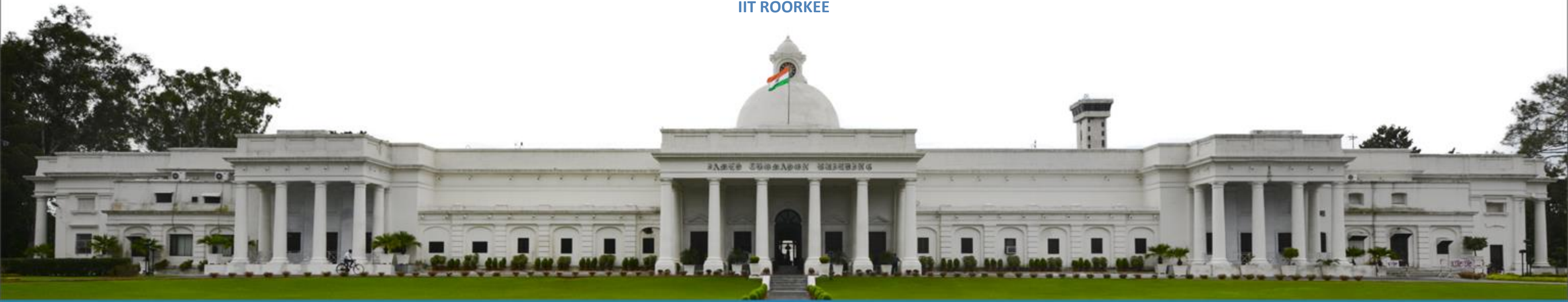
Data Analytics with Python

Lecture 9: Probability Distributions-II

Dr. A. Ramesh

DEPARTMENT OF MANAGEMENT

IIT ROORKEE



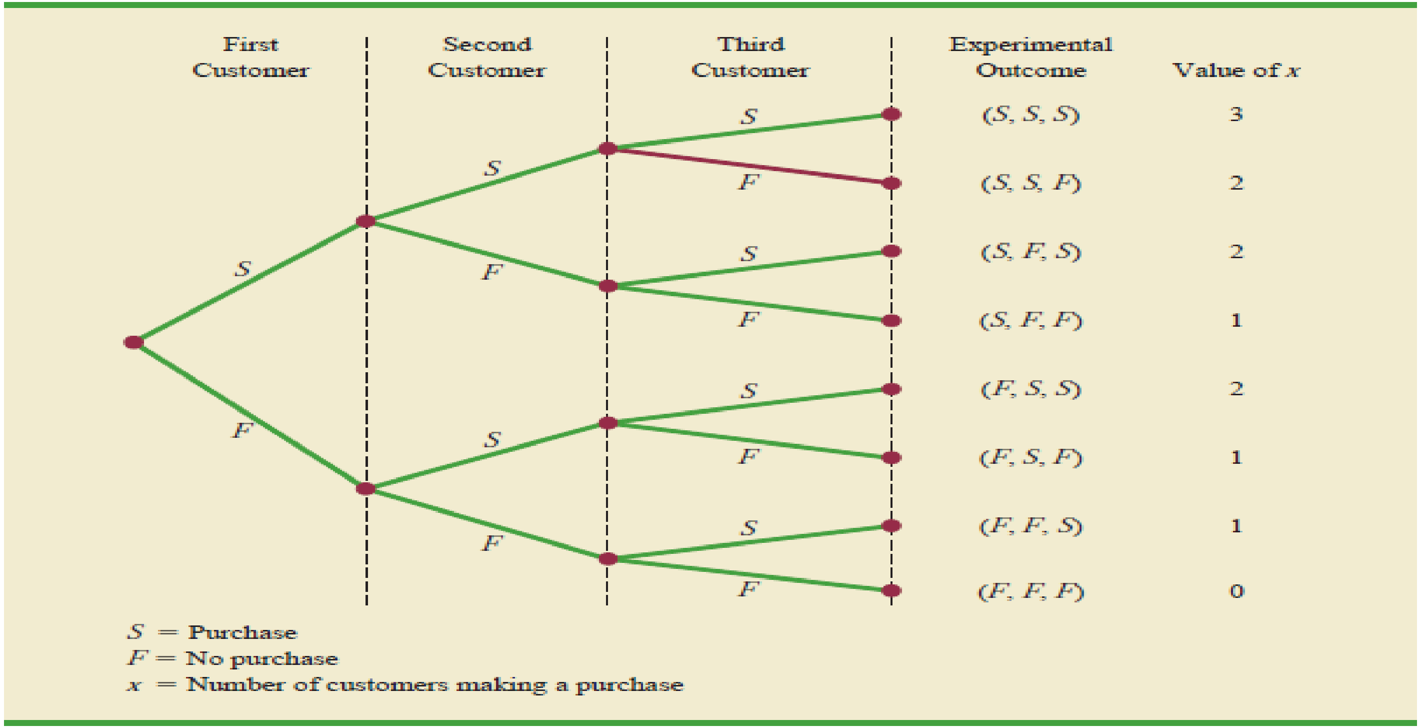
Some Special Distributions

- Discrete
  - Binomial
  - Poisson
  - Hyper geometric
- Continuous
  - Uniform
  - Exponential
  - Normal

Binomial Distribution

- Let us consider the purchase decisions of the next three customers who enter a store.
- On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is .30.
- What is the probability that two of the next three customers will make a purchase?

Tree diagram for the Martin clothing store problem

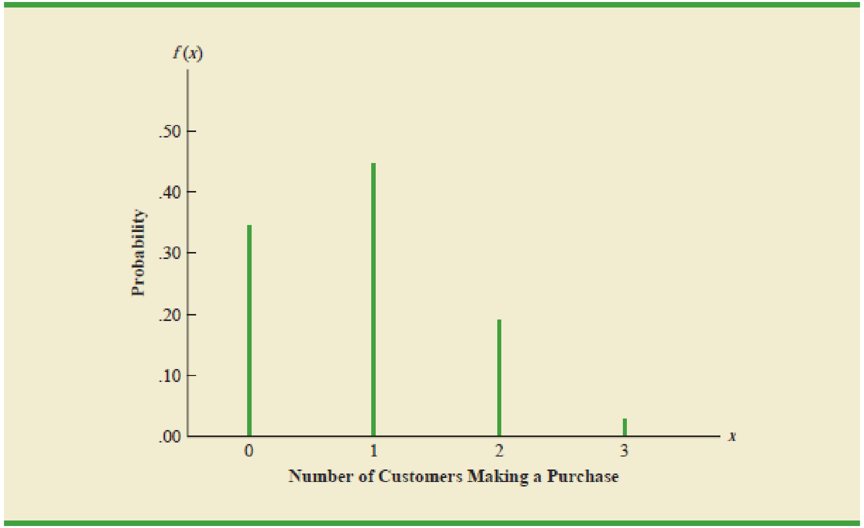


Trial Outcomes

Trial Outcomes				
1st Customer	2nd Customer	3rd Customer	Experimental Outcome	Probability of Experimental Outcome
Purchase	Purchase	No purchase	(S, S, F)	$pp(1-p) = p^2(1-p) = (.30)^2(.70) = .063$
Purchase	No purchase	Purchase	(S, F, S)	$p(1-p)p = p^2(1-p) = (.30)^2(.70) = .063$
No purchase	Purchase	Purchase	(F, S, S)	$(1-p)pp = p^2(1-p) = (.30)^2(.70) = .063$

Graphical representation of the probability distribution for the number of customers making a purchase

x	P(x)
0	0.7 x 0.7 x 0.7=0.343
1	0.3x0.7x07+ 0.7x0.3x0.7+ 0.7x0.7x0.3 = 0.441
2	0.189
3	0.027



Binomial Distribution- Assumtions

- Experiment involves n identical trials
- Each trial has exactly two possible outcomes: success and failure
- Each trial is independent of the previous trials
- p is the probability of a success on any one trial

  q = (1-p) is the probability of a failure on any one trial
- p and q are constant throughout the experiment
- X is the number of successes in the n trials

Binomial Distribution

- Probability function

$$P(X) = \frac{n!}{X!(n-X)!} p^X \cdot q^{n-X} \quad \text{for} \quad 0 \le X \le n$$

- Mean value

$$\boxed{\mu = n \cdot p}$$

- Variance and standard deviation

$$\sigma^2 = n \cdot p \cdot q$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{n \cdot p \cdot q}$$

Binomial Table

SELECTED VALUES FROM THE BINOMIAL PROBABILITY TABLE

EXAMPLE: $n = 10$, $x = 3$, $p = .40$; $f(3) = .2150$

		p									
n	*x*	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	**.2150**	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010

Mean and Variance

- Suppose that for the next month the Clothing Store forecasts 1000 customers will enter the store.
- What is the expected number of customers who will make a purchase?
- The answer is $\mu = np =$ (1000)(.3) = 300.
- For the next 1000 customers entering the store, the variance and standard deviation for the number of customers who will make a purchase are

$$\sigma^2 = np(1-p) = 1000(.3)(.7) = 210$$
$$\sigma = \sqrt{210} = 14.49$$

Poisson Distribution

- Describes discrete occurrences over a continuum or interval
- A discrete distribution
- Describes rare events
- Each occurrence is independent any other occurrences.
- The number of occurrences in each interval can vary from zero to infinity.
- The expected number of occurrences must hold constant throughout the experiment.

Poisson Distribution: Applications

- **Arrivals at queuing systems**
  - airports -- people, airplanes, automobiles, baggage
  - banks -- people, automobiles, loan applications
  - computer file servers -- read and write operations

- **Defects in manufactured goods**
  - number of defects per 1,000 feet of extruded copper wire
  - number of blemishes per square foot of painted surface
  - number of errors per typed page

Poisson Distribution

- Probability function

$$P(X) = \frac{\lambda^{X} e^{-\lambda}}{X!} \quad \text{for} \quad X = 0, 1, 2, 3, \ldots$$

$where:$

$\lambda = long - run\ average$

$e = 2.718282\ldots \quad (\text{the base of natural logarithms})$

Mean value	Variance	Standard deviation
$\boxed{\lambda}$	$\boxed{\lambda}$	$\boxed{\sqrt{\lambda}}$

Poisson Distribution: Example

$\lambda = 3.2$ customers/4 minutes

X = 10 customers/8 minutes

Adjusted λ

λ=6.4 customers/8 minutes

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

$$P(X=10) = \frac{6.4^{10} e^{-6.4}}{10!} = 0.0528$$

$\lambda = 3.2$ customers/4 minutes

X = 6 customers/8 minutes

Adjusted λ

λ=6.4 customers/8 minutes

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

$$P(X=6) = \frac{6.4^{6} e^{-6.4}}{6!} = 0.1586$$

Poisson Probability Table

Example: μ = 10, x = 5; f (5) = .0378

	μ									
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	**.0378**
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

The Hypergeometric Distribution

- The **binomial distribution** is applicable when selecting from a finite population with replacement or from an infinite population without replacement.

- The hypergeometric distribution is applicable when selecting from a finite population without replacement.

Hyper Geometric Distribution

- Sampling without replacement from a finite population
- The number of objects in the population is denoted N.
- Each trial has exactly two possible outcomes, success and failure.
- Trials are not independent
- X is the number of successes in the n trials
- The binomial is an acceptable approximation, if $N/10 > n$ Otherwise it is not.

Hypergeometric Distribution

- Probability function
  - N is population size
  - n is sample size
  - A is number of successes in population
  - x is number of successes in sample

$$P(x) = \frac{\left({}_{A}C_{x}\right)\left({}_{N-A}C_{n-x}\right)}{{}_{N}C_{n}}$$

- Mean Value

$$\mu = \frac{A \cdot n}{N}$$

- Variance and standard deviation

$$\sigma^2 = \frac{A(N-A)n(N-n)}{N^2(N-1)}$$

$$\sigma = \sqrt{\sigma^2}$$

The Hypergeometric Distribution Example

- Different computers are checked from 10 in the department. 4 of the 10 computers have illegal software loaded.
- What is the probability that 2 of the 3 selected computers have illegal software loaded?
- So, N = 10, n = 3, A = 4, X = 2

$$P(X=2)=\frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}}=\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}=\frac{(6)(6)}{120}=0.3$$

- The probability that 2 of the 3 selected computers have illegal software loaded is .30, or 30%.

Continuous Probability Distributions

- A continuous random variable is a variable that can assume any value on a continuum (can assume an uncountable number of values)
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height
- These can potentially take on any value, depending only on the ability to measure precisely and accurately.

Continuous Distributions

- Uniform
- Normal
- Exponential

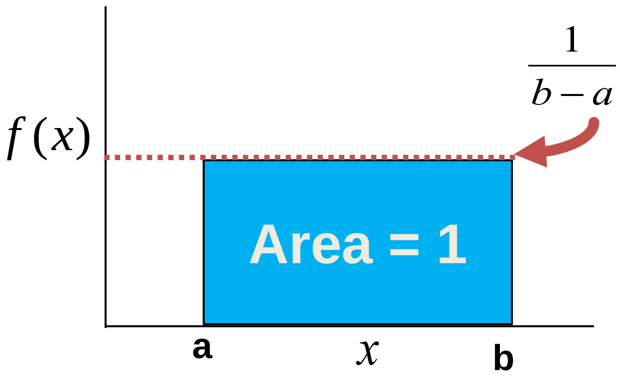
The Uniform Distribution

- The uniform distribution is a probability distribution that has equal probabilities for all possible outcomes of the random variable

- Because of its shape it is also called a rectangular distribution

Uniform Distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a} & for \quad a \le x \le b \\ 0 & for \quad \text{all other values} \end{cases}$$



Uniform Distribution: Mean and Standard Deviation

Mean

$$\mu = \frac{a+b}{2}$$

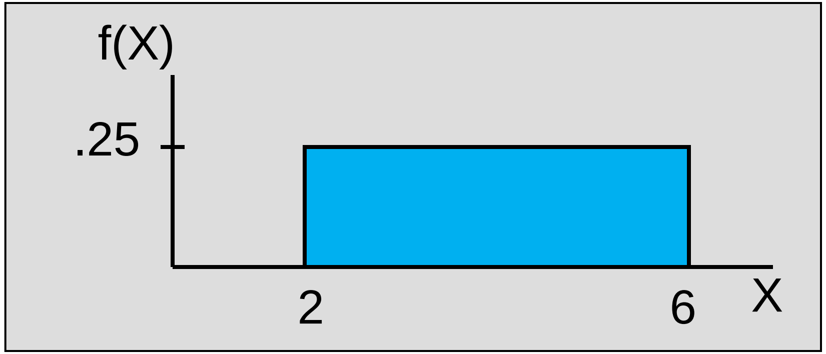
Standard Deviation

$$\sigma = \frac{b-a}{\sqrt{12}}$$

The Uniform Distribution

Example: Uniform probability distribution over the range $2 \le X \le 6$:

$$f(X) = \frac{1}{6-2} = .25 \quad \text{for } 2 \le X \le 6$$

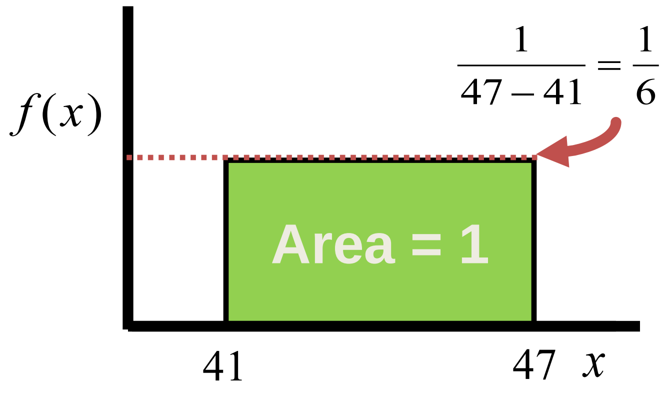


$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(6-2)^2}{12}} = 1.1547$$

Uniform Distribution Example

$$f(x) = \begin{cases} \dfrac{1}{47-41} & \textit{for} \quad 41 \le x \le 47 \\ 0 & \textit{for} \quad \text{all other values} \end{cases}$$



Uniform Distribution: Mean and Standard Deviation

Mean

$$\mu = \frac{a+b}{2}$$

Mean

$$\mu = \frac{41+47}{2} = \frac{88}{2} = 44$$

Standard Deviation

$$\sigma = \frac{b-a}{\sqrt{12}}$$

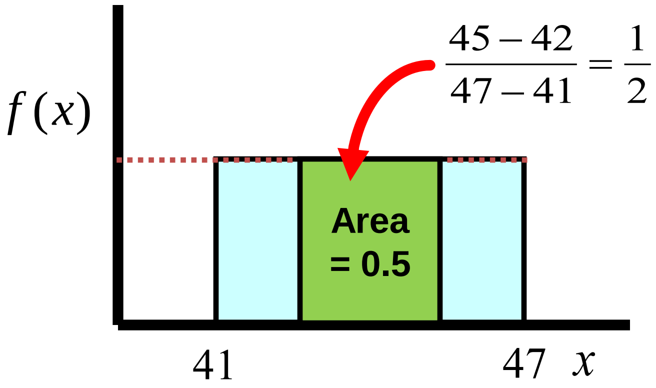
Standard Deviation

$$\sigma = \frac{47-41}{\sqrt{12}} = \frac{6}{3.464} = 1.732$$

Uniform Distribution Probability

$$P(x_1 \le X \le x_2) = \frac{x_2 - x_1}{b - a}$$

$$P(42 \le X \le 45) = \frac{45-42}{47-41} = \frac{1}{2}$$



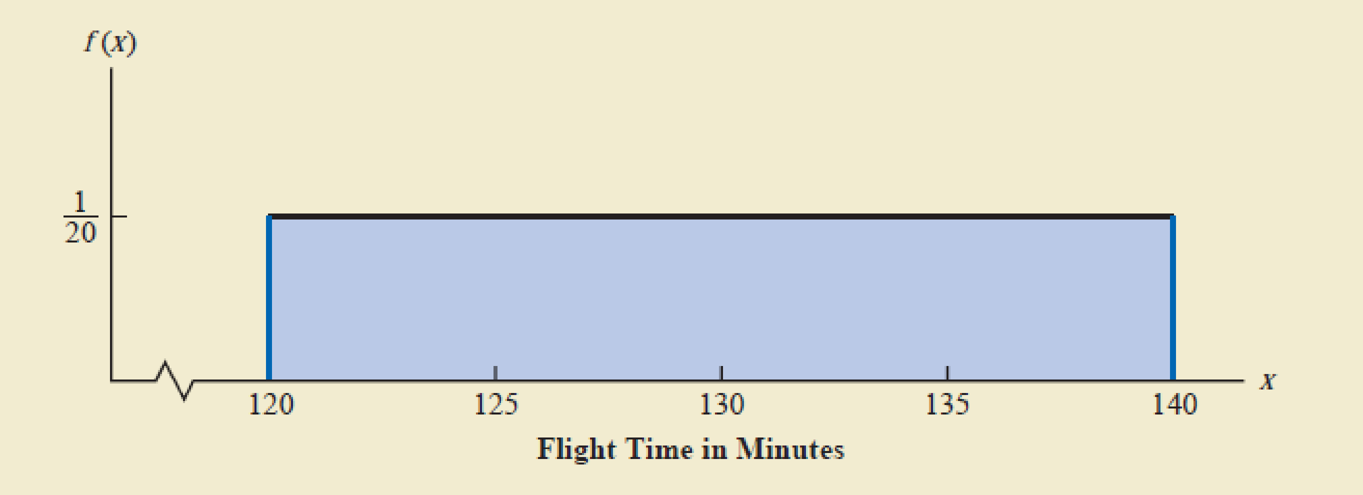
Example : Uniform Distribution

- Consider the random variable x representing the flight time of an airplane traveling from Delhi to Mumbai.
- Suppose the flight time can be any value in the interval from 120 minutes to 140 minutes.
- Because the random variable x can assume any value in that interval, x is a continuous rather than a discrete random variable

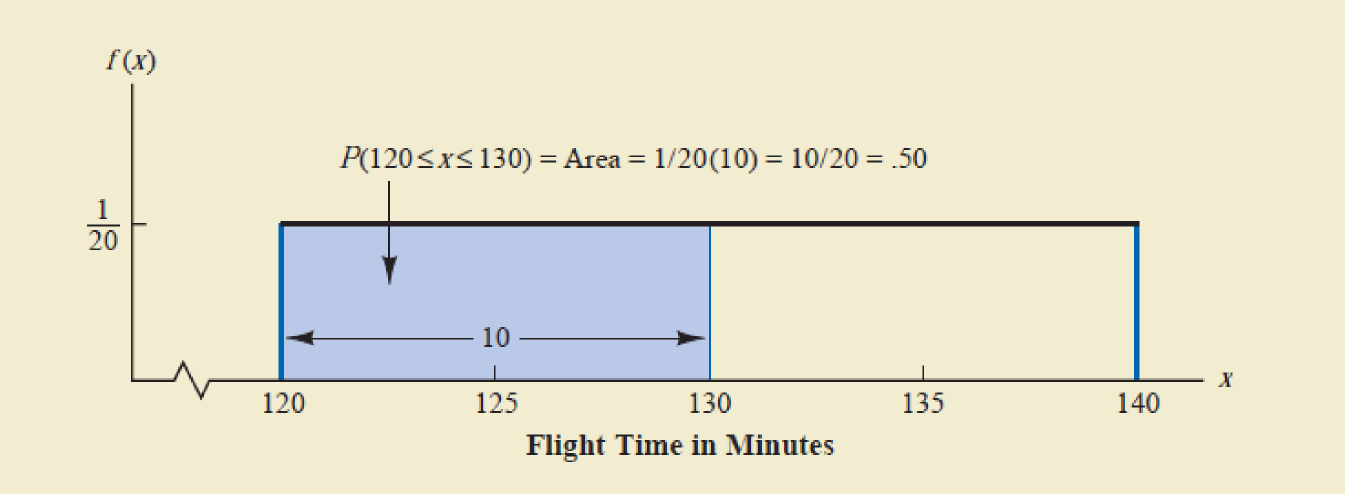
Example : Uniform Distribution contd….

- Let us assume that sufficient actual flight data are available to conclude that the probability of a flight time within any 1-minute interval is the same as the probability of a flight time within any other 1-minute interval contained in the larger interval from 120 to 140 minutes.
- With every 1-minute interval being equally likely, the random variable x is said to have a uniform probability distribution.

Uniform Probability Distribution for Flight time



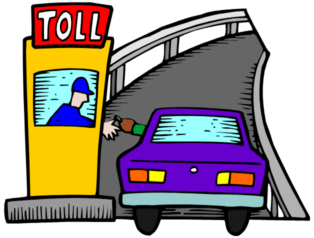
Probability of a flight time between 120 and 130 minutes



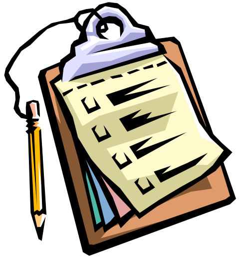
Exponential Probability Distribution

- The exponential probability distribution is useful in describing the time it takes to complete a task.
- The exponential random variables can be used to describe:

Time between vehicle arrivals at a toll booth



Time required to complete a questionnaire



Distance between major defects in a highway

Exponential Probability Distribution

- Density Function

$$f(x) = \frac{1}{\mu} e^{-x/\mu}$$

where: μ = mean

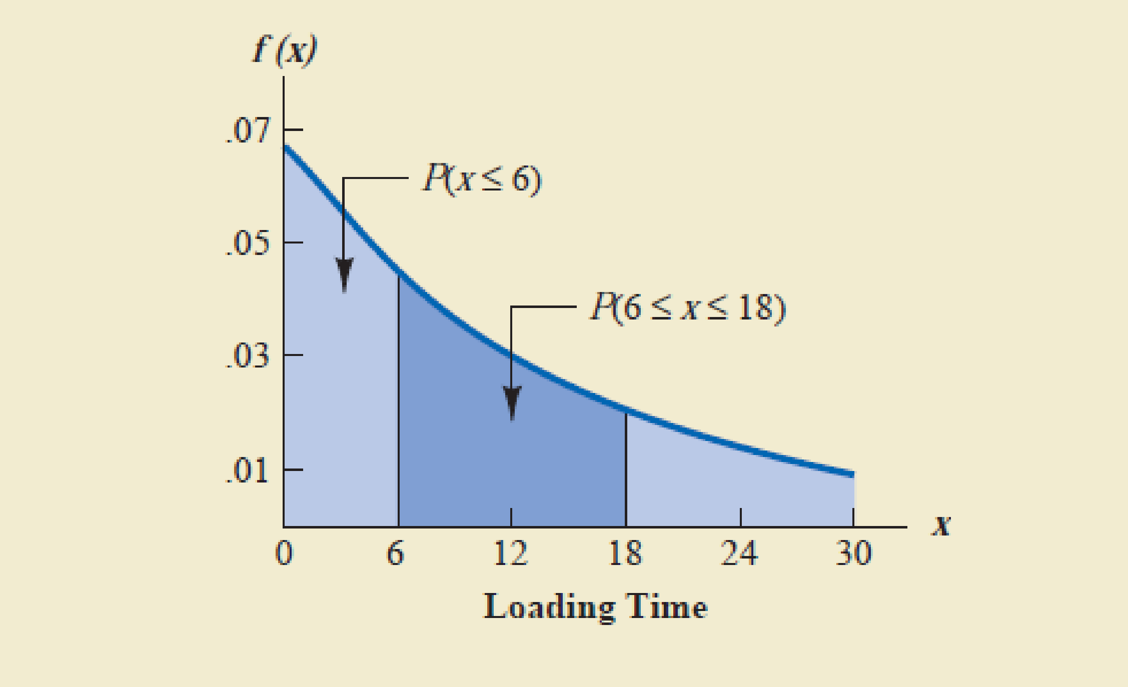
e = 2.71828

Exponential Probability Distribution

- Suppose that *x* represents the loading time for a truck at loading dock and follows such a distribution.
- If the mean, or average, loading time is 15 minutes (μ = 15), the appropriate probability density function for *x* is

$$f(x) = \frac{1}{15} e^{-x/15}$$

Exponential Distribution for the loading Dock Example



Exponential Probability Distribution
Cumulative Probabilities

$$P(x \le x_0) = 1 - e^{-x_0/\mu}$$

where:

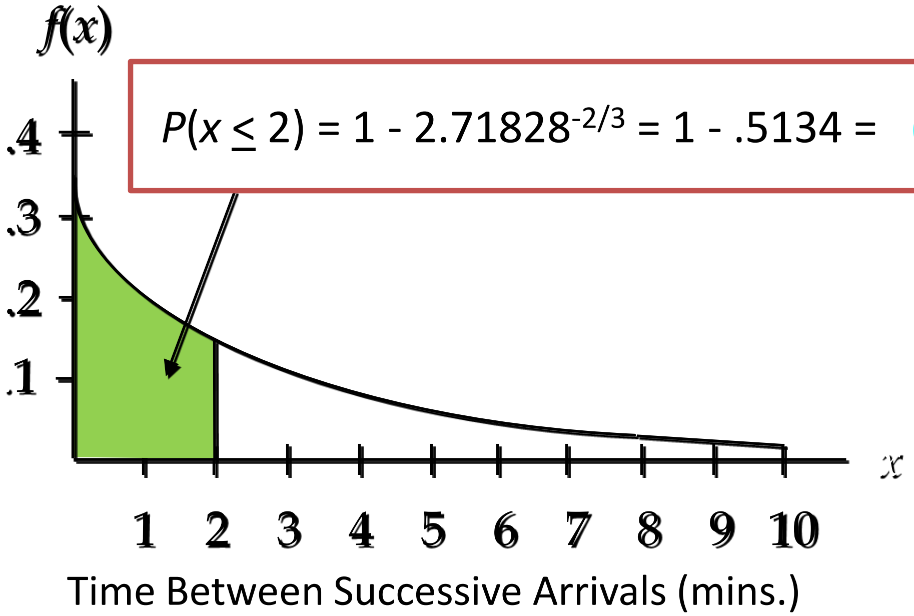
x0 = some specific value of x

Example: Exponential Probability Distribution

- The time between arrivals of cars at a Petrol pump follows an exponential probability distribution with a mean time between arrivals of 3 minutes.
- The Petrol pump owner would like to know the probability that the time between two successive arrivals will be 2 minutes or less.

Example: Petrol Pump Problem

$P(x \leq 2) = 1 - 2.71828^{-2/3} = 1 - .5134 = 4866$



Relationship between the Poisson and Exponential Distributions

The Poisson distribution provides an appropriate description of the number of occurrences per interval

The exponential distribution provides an appropriate description of the length of the interval between occurrences

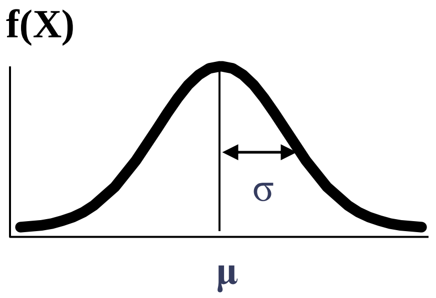
Mean of Poisson and Mean of Exponential Distributions

- Because the average number of arrivals is 10 cars per hour, the average time between cars arriving is

$$\frac{1 \text{ hour}}{10 \text{ cars}} = .1 \text{ hour/car}$$

The Normal Distribution: Properties

- ‘Bell Shaped’
- Symmetrical
- Mean, Median and Mode are equal
- Location is characterized by the mean, μ
- Spread is characterized by the standard deviation, σ
- The random variable has an infinite theoretical range: -∞ to +∞



Mean = Median = Mode

The Normal Distribution: Density Function

The formula for the normal probability density function is

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{(X-\mu)}{\sigma}\right)^2}$$

Where e = the mathematical constant approximated by 2.71828

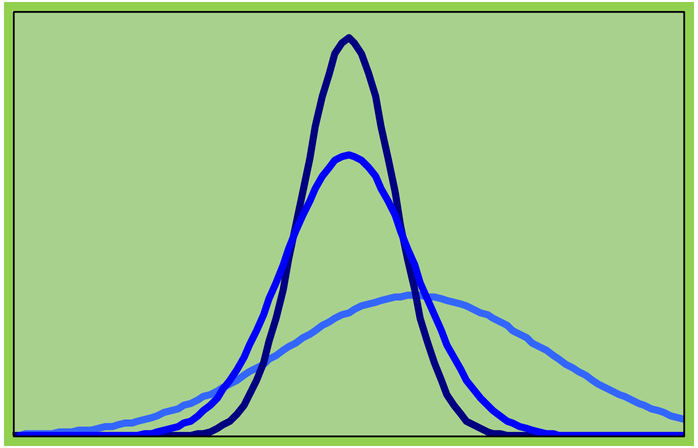
π = the mathematical constant approximated by 3.14159

μ = the population mean

σ = the population standard deviation

X = any value of the continuous variable

The Normal Distribution: Shape



By varying the parameters μ and σ, we obtain different normal distributions