

IIT KHARAGPUR



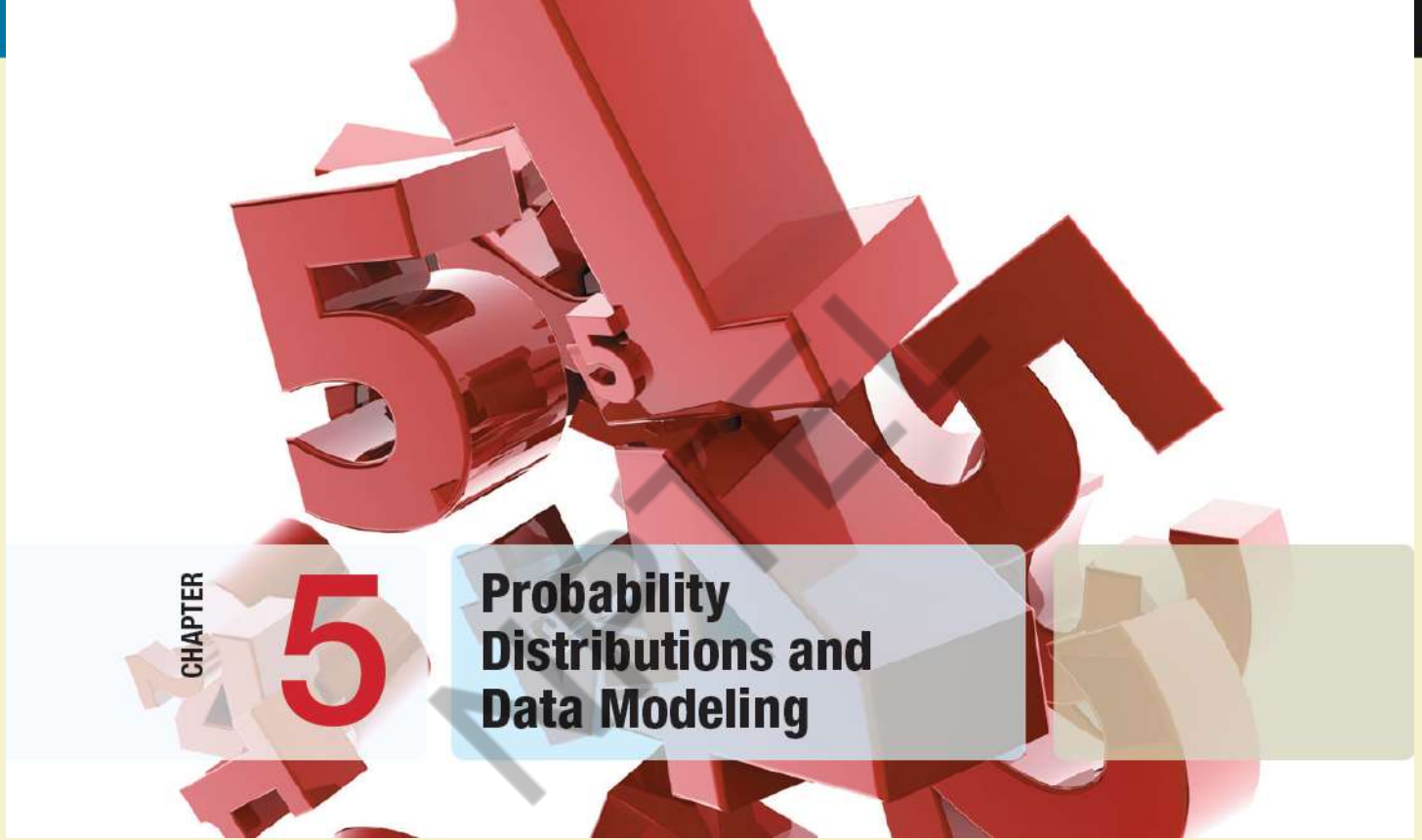
NPTEL ONLINE
CERTIFICATION COURSES

Business Analytics for Management Decision

Probability Distributions and Data Modeling

Rudra P. Pradhan
Vinod Gupta School of Management
IIT KHARAGPUR

NPTEL



CHAPTER 5

Probability Distributions and Data Modeling

Chapter 5 Topics

- Basic Concepts of Probability
- Random Variables and Probability Distributions
- Discrete Probability Distributions
- Continuous Probability Distributions
- Random Sampling from Probability Distributions
- Data Modeling and Distribution Fitting

NPTEL

Basic Concepts of Probability

- Probability is the likelihood that an outcome occurs.
- An experiment is the process that results in an outcome.
- The outcome of an experiment is a result that we observe.
- The sample space is the collection of all possible outcomes of an experiment.
- An event is a collection of one or more outcomes from a sample space.

NPTEL

Basic Concepts of Probability

Probabilities may be defined from one of three perspectives:

- Classical definition: probabilities can be deduced from theoretical arguments
- Relative frequency definition: probabilities are based on empirical data
- Subjective definition: probabilities are based on judgment and experience

NPTEL

Basic Concepts of Probability

Example 5.1 Classical Definition of Probability

Suppose we roll 2 dice

- Probability die rolls sum to three = 2/36

Suppose two consumers try a new product.

- Assume equally likely possible outcomes:
  1. like, like
  2. like, dislike
  3. dislike, like
  4. dislike, dislike
- Probability at least one dislikes product = 3/4

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

NPTEL

Basic Concepts of Probability

Example 5.2

Relative Frequency Definition of Probability

- Probability a computer is repaired in 10 days

  = 0.076

	A	B	C	D
1	Computer Repair Times			
2				
3			Relative	Cumulative
4	Days	Frequency	Frequency	Percentage
5	0	0	0.000	0.0%
6	1	0	0.000	0.0%
7	2	0	0.000	0.0%
8	3	0	0.000	0.0%
9	4	0	0.000	0.0%
10	5	1	0.004	0.4%
11	6	2	0.008	1.2%
12	7	5	0.020	3.2%
13	8	12	0.048	8.0%
14	9	14	0.056	13.6%
15	10	19	0.076	21.2%
16	11	19	0.076	28.8%
17	12	23	0.092	38.0%
18	13	22	0.088	46.8%
19	14	20	0.080	54.8%
20	15	23	0.092	64.0%
21	16	18	0.072	71.2%
22	17	16	0.064	77.6%
23	18	9	0.036	81.2%
24	19	9	0.036	84.8%
25	20	8	0.032	88.0%
26	21	5	0.020	90.0%



NPTEL

Figure 5.1

Basic Concepts of Probability

Two Basic Facts Govern Probability

- The probability associated with any outcome must be between 0 and 1.

$$0 \leq P(O_i) \leq 1 \quad \text{for each outcome } O_i$$

- The sum of the probabilities over all possible outcomes must be equal to 1.

$$P(O_1) + P(O_2) + \ldots + P(O_n) = 1$$

(n is the number of outcomes in the sample space.)

NPTEL

Basic Concepts of Probability

Example 5.3

Computing the Probability of an Event

Consider the events:

- Rolling 7 or 11 on two dice

  Probability = 6/36 + 2/36 = 8/36.

- Repair a computer in 7 days or less

  Probability =

  $= O_1 + O_2 + O_3 + O_4 + O_5 + O_6 + O_7$

  = 0 + 0 + 0 + 0 + .004 + .008 + .002

  = 0.032

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

	A	B	C
1	Computer Repair Times		
2			
3			Relative
4	Days	Frequency	Frequency
5	0	0	0.000
6	1	0	0.000
7	2	0	0.000
8	3	0	0.000
9	4	0	0.000
10	5	1	0.004
11	6	2	0.008
12	7	5	0.020

From Figure 5.1

NPTEL

Basic Concepts of Probability

Example 5.4 Computing the Probability of the Complement of an Event $P(A^c) = 1 - P(A)$

- A^c, the complement of A, consists of all outcomes in the sample space not in A.

Dice example:

- $A = \{7, 11\}$

  $P(A) = 8/36$

- $A^c = \{2, 3, 4, 5, 6, 8, 9, 10, 12\}$

  $P(A^c) = 1 - 8/36 = 28/36$

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

NPTEL

Basic Concepts of Probability

Example 5.5 Computing the Probability of Mutually Exclusive Events

$$P(A \text{ or } B) = P(A) + P(B)$$

- Mutually exclusive events have no outcomes in common.

Dice Example:

- $A = \{7, 11\}$
- $B = \{2, 3, 12\}$
- $P(A \text{ or } B)$ = UNION of events A and B

$= P(A) + P(B)$

$= 8/36 + 4/36 = 12/36$

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

NPTEL

Basic Concepts of Probability

Example 5.6 Computing the Probability of Non-Mutually Exclusive Events

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Dice Example:

- A = {2, 3, 12}
- B = {even number}
- $P(A \text{ or } B)$ = UNION of events A and B

$$= P(A) + P(B) - P(A \text{ and } B)$$
$$= 4/36 + 18/36 - 2/36$$
$$= 20/36$$

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

NPTEL

Basic Concepts of Probability

Example 5.7 Conditional Probability in Marketing

- The Data shows the first and second purchases for a sample of 200 customers.

	A	B
1	Apple Purchase History	
2		
3	First Purchase	Second Purchase
4	iPod	iMac
5	iPhone	MacBook
6	iMac	iPhone
7	iPhone	iPod
8	iPod	iPhone
9	MacBook	iPod
10	iPhone	MacBook

Figure 5.2

- Probability of purchasing an iPad given already purchased an iMac = 2/13

Count of Second Purchase	Colu					
Row Labels	iMac	iPad	iPhone	iPod	MacBook	Grand Total
iMac		2	3	2	6	13
iPad	1		1	2	10	14
iPhone	3	4		11	21	42
iPod	3	12	12		30	57
MacBook	8	16	26	24		74
Grand Total	15	34	42	42	67	200

Figure 5.3

NPTEL

Basic Concepts of Probability

Example 5.8 Computing a Conditional Probability in a Cross-Tabulation

- Probability of preferring Brand 1 given that a respondent is male = 25/63

	A	B	C	D	E	F	G	H	I
1	Energy Drink Survey								
2									
3	Respondent	Gender	Brand Preference						
4	1	Male	Brand 3		Count of Respondent	Column Labels			
5	2	Female	Brand 3		Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		Grand Total	34	23	43	100
9	6	Female	Brand 2						
10	7	Male	Brand 2						
11	8	Female	Brand 2						
12	9	Male	Brand 1						
13	10	Female	Brand 3						

Figure 5.4

NPTEL

Basic Concepts of Probability

Example 5.9

Using the Conditional Probability Formula

- Probability of *A* given *B*:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

- $P(B_1|M) = P(B_1 \text{ and } M)/ P(M)$
  $= (25/100)/(63/100)$
  $= 25/63 = 0.397$

Summary of conditional probabilities:

P(Brand|Gender)	Brand 1	Brand 2	Brand 3
Male	0.397	0.270	0.333
Female	0.243	0.162	0.595

NPTEL

Basic Concepts of Probability

Example 5.10

Using the Multiplication Law of Probability

$$P(A \text{ and } B) = P(A|B)\,P(B) = P(B|A)\,P(A)$$

Texas Hold 'Em Poker Game

- Probability of pocket aces (two aces in hand):
- *P*(Ace on first card and Ace on second card)

  = $P(A_1 \text{ and } A_2)$

  = $P(A_2|A_1)\ P(A_1)$

  = (3/51) (4/52)

  = 0.004525

NPTEL

Basic Concepts of Probability

Example 5.11 Determining if Two Events are Independent

$$P(A|B) = P(A)$$

- Are Gender and Brand Preference Independent?

Count of Respondent	Column Labels			
Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
Female	9	6	22	37
Male	25	17	21	63
Grand Total	34	23	43	100

P(Brand|Gender)	Brand 1	Brand 2	Brand 3
Male	0.397	0.270	0.333
Female	0.243	0.162	0.595

- Is $P(B_1|M) = P(B_1)$?

  $0.397 \neq .34$

  Gender and Brand Preference are Dependent.

NPTEL

Basic Concepts of Probability

Example 5.12

Using the Multiplication Law for Independent Events

$$P(A \text{ and } B) = P(B)\,P(A) = P(A)P(B)$$

- Dice Roll Example:
- Rolling pairs of dice are independent events since they do not depend on the previous rolls.
- *A* = {roll a sum of 6 on first pair die rolls}
- *B* = {roll a sum of 2, 3, or 12 on second pair rolls}
- *P(A* and *B) = P(A) P(B)*

  = (5/36) (4/36) = 0.0154

NPTEL

Random Variables and Probability Distributions

- A random variable is a numerical description of the outcome of an experiment.
- A discrete random variable is one for which the number of possible outcomes can be counted.
- A continuous random variable has outcomes over one or more continuous intervals of real numbers.
- A probability distribution is a characterization of the possible values that a random variable may assume along with the probability of assuming these values.

NPTEL

Random Variables and Probability Distributions

Example 5.13

Discrete and Continuous Random Variables

Examples of Discrete Variables:

- outcomes of dice rolls
- whether a customer likes or dislikes a product
- number of hits on a Web site link today

Examples of Continuous Variables:

- weekly change in DJIA
- daily temperature
- time between machine failures

NPTEL

Random Variables and Probability Distributions

Example 5.14

Probability Distribution of Dice Rolls

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Outcome	2	3	4	5	6	7	8	9	10	11	12
3	Probability	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

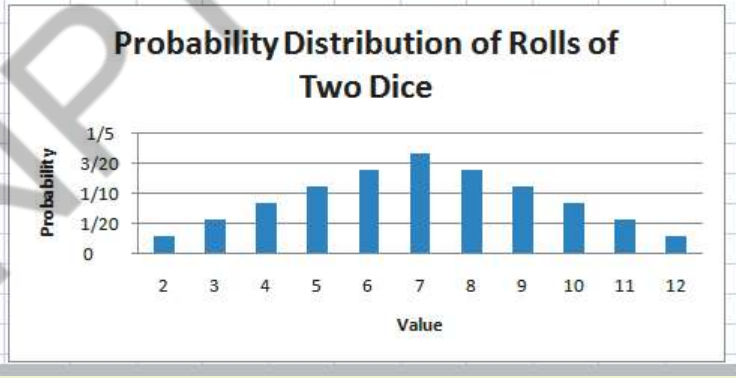


Figure 5.5

NPTEL

Random Variables and Probability Distributions

Empirical Probability Distribution - Sample data is used to approximate a probability distribution

Computer Repair Times				
Days	Frequency	Relative Frequency	Cumulative Percentage	
0	0	0.000	0.0%	
1	0	0.000	0.0%	
2	0	0.000	0.0%	
3	0	0.000	0.0%	
4	0	0.000	0.0%	
5	1	0.004	0.4%	
6	2	0.008	1.2%	
7	5	0.020	3.2%	
8	12	0.048	8.0%	
9	14	0.056	13.6%	
10	19	0.076	21.2%	
11	19	0.076	28.8%	
12	23	0.092	38.0%	
13	22	0.088	46.8%	
14	20	0.080	54.8%	
15	23	0.092	64.0%	
16	18	0.072	71.2%	
17	16	0.064	77.6%	
18	9	0.036	81.2%	
19	9	0.036	84.8%	
20	8	0.032	88.0%	
21	5	0.020	90.0%	



Figure 5.1

NPTEL

Random Variables and Probability Distributions

Example 5.15 A Subjective Probability Distribution

- Distribution of an expert's assessment of how the DJIA might change next year.

	A	B
1	Change in DJIA	Subjective Probability
2	-20%	0.01
3	-15%	0.05
4	-10%	0.08
5	-5%	0.15
6	0%	0.2
7	5%	0.25
8	10%	0.18
9	15%	0.06
10	20%	0.02
11		

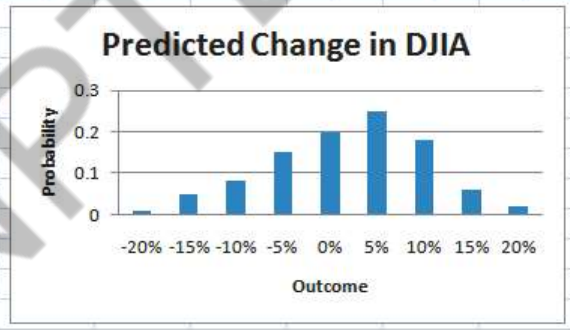


Figure 5.6

NPTEL

Discrete Probability Distributions

Probability Mass Function

a mathematical function *f(x)* specifying the probability of the random variable *X.*

x_i represents the *i* th value of *X.*

Properties:

$$0 \leq f(x_i) \leq 1 \qquad \text{for all } i$$

$$\sum_i f(x_i) = 1$$

Cumulative distribution function:

$F(x) = P(X \leq x)$

NPTEL

Discrete Probability Distributions

Example 5.16

Probability Mass Function for Rolling Two Dice

$f(x_2) = 1/36$

$f(x_3) = 2/36$

$f(x_4) = 3/36$

$f(x_5) = 4/36$

$f(x_6) = 5/36$

:

$f(x_{12}) = 1/36$

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Outcome	2	3	4	5	6	7	8	9	10	11	12
3	Probability	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

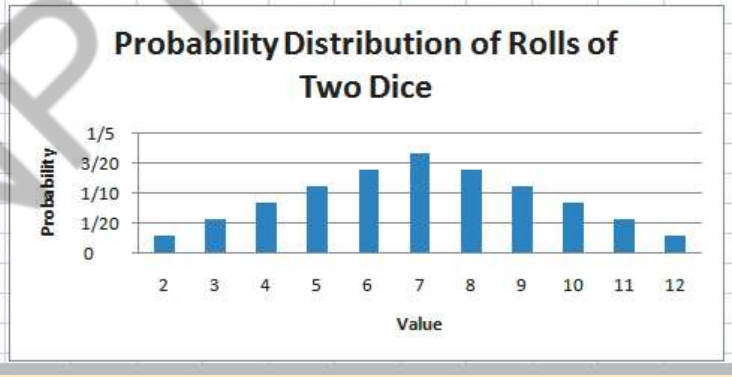


Figure 5.5

NPTEL

Discrete Probability Distributions

Example 5.17

Using the Cumulative Distribution Function

- Probability of rolling between 4 and 8:

$= P(4 \le X \le 8)$

$= P(3 < X \le 8)$

$= F(x_8) - F(x_3)$

$=13/18 - 1/12$

$= 23/36$

	A	B	C	D	E	F	G	H	I	J	K	L
18	Outcome	2	3	4	5	6	7	8	9	10	11	12
19	Probability	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36
20	Cumulative Probability	1/36	1/12	1/6	5/18	5/12	7/12	13/18	5/6	11/12	35/36	1

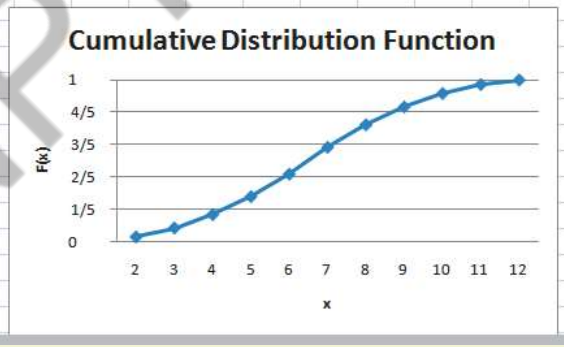


Figure 5.7

NPTEL

Discrete Probability Distributions

Example 5.18 Computing the Expected Value of the sum of values on 2 die rolls

$$E[X] = \sum_{i=1}^{\infty} x_i f(x_i)$$

E[X] = 2(1/36) + 3(1/18) + …

12(1/36) = 7

	A	B	C
1	Expected Value Calculations		
2			
3	Outcome, x	Probability, f(x)	x*f(x)
4	2	1/36	1/18
5	3	1/18	1/6
6	4	1/12	1/3
7	5	1/9	5/9
8	6	5/36	5/6
9	7	1/6	1 1/6
10	8	5/36	1 1/9
11	9	1/9	1
12	10	1/12	5/6
13	11	1/18	11/18
14	12	1/36	1/3
15		Expected value	7

Figure 5.8

NPTEL

Discrete Probability Distributions

Example 5.19 Expected Value on Television

Apprentice example

- Teams were required to select an artist (mainstream or avant-garde) and sell their art for the most money possible.

Deal or No Deal example

- Contestant had 5 briefcases left with $100, $400, $1000, $50,000 or $300,000 in them.
- Expected value of briefcases is $70,300.
- Banker offered contestant $80,000 to quit.

NPTEL

Discrete Probability Distributions

Example 5.20 Expected Value of Charitable Raffle

- Cost of raffle ticket is $50
- 1000 raffle tickets were sold.
- Prize for winning raffle is $25,000

x	$f(x)$
−$50	0.999
$24,950	0.001

$E[X] = -\$25$

NPTEL

Discrete Probability Distributions

Example 5.21 Airline Revenue Management

- Full and discount airfares are available for a flight.
- Full-fare ticket costs $560
- Discount ticket costs $400
- X = ticket price paid
- p = 0.75 (the probability of selling a full-fare ticket)
- $E[X]$ = 0.75($560) + 0.25(0) = $420
- The airline should not discount full-fare tickets because the expected value of a full-fare ticket is greater than the cost of a discount ticket.
- Break-even point: $400 = p($560) or p = 0.714

NPTEL

Discrete Probability Distributions

Example 5.22

Computing the Variance of a Random Variable

$$\text{Var}[X] = \sum_{j=1}^{\infty} (x_j - E[X])^2 f(x_j)$$

	A	B	C	D	E	F
1	Variance Calculations					
2						
3	Outcome, x	Probability, f(x)	x*f(x)	(x - E[X])	(x - E[X])^2	(x - E[X])^2*f(x)
4	2	1/36	1/18	-5	25	25/36
5	3	1/18	1/6	-4	16	8/9
6	4	1/12	1/3	-3	9	3/4
7	5	1/9	5/9	-2	4	4/9
8	6	5/36	5/6	-1	1	5/36
9	7	1/6	1 1/6	0	0	0
10	8	5/36	1 1/9	1	1	5/36
11	9	1/9	1	2	4	4/9
12	10	1/12	5/6	3	9	3/4
13	11	1/18	11/18	4	16	8/9
14	12	1/36	1/3	5	25	25/36
15		Expected value	7		Variance	5 5/6

Figure 5.9

NPTEL

Discrete Probability Distributions

Bernoulli Distribution

- two possible outcomes each with a constant probability of occurrence
- typically "success" is $x = 1$ and "failure" $x = 0$
- p is the probability of a success outcome

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

$E[X] = p$

$\text{Var}[X] = p(1 - p)$

NPTEL

Discrete Probability Distributions

Example 5.23 Using the Bernoulli Distribution

Model whether an individual responds positively to a telemarketing promotion.

- You have a box with 20 red and 80 white marbles.
- You ask individuals exposed to the telemarketing promotion to select a marble and then replace it.
- If the customer selects a red marble, the customer makes a purchase.
- If the customer selects a white marble, the customer does not make a purchase.

NPTEL

Discrete Probability Distributions

Binomial Distribution

- Models *n* independent replications of a Bernoulli experiment
- *X* represents the number of successes in these *n* experiments

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{for } x = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

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

NPTEL

Discrete Probability Distributions

Example 5.24 Computing Binomial Probabilities

- Suppose 10 individuals receive the telemarking promotion.
- Each individual has a 0.2 probability of making a purchase.
- Find the probability that exactly 3 of the 10 individuals make a purchase.

$$f(3) = \binom{10}{3}(0.2)^3(0.8)^{10-3}$$

$$= 10!/3!7!\ (0.008)(0.2097152)$$

$$= 120(0.008)(0.2097152) = 0.20133$$

NPTEL

Discrete Probability Distributions

Example 5.25

Using Excel's Binomial Distribution Function

BINOM.DIST(*number_s*, *trials*, *probability_s*, *cumulative*)

True: *F(x)*

False: *f(x)*

$P(x = 3) = 0.20133$

$= f(3)$

=BINOM.DIST(3, 10, 0.2, true)

$P(x \leq 3) = 0.87913$

$= F(3)$

=BINOM.DIST(3, 10, 0.2, false)

	A	B	C
1	Binomial Probabilities		
2			
3	n	10	
4	p	0.2	
5			
6	x	f(x)	F(x)
7	0	0.10737	0.10737
8	1	0.26844	0.37581
9	2	0.30199	0.67780
10	3	0.20133	0.87913
11	4	0.08808	0.96721
12	5	0.02642	0.99363
13	6	0.00551	0.99914
14	7	0.00079	0.99992
15	8	0.00007	1.00000
16	9	0.00000	1.00000
17	10	0.00000	1.00000

Figure 5.10

NPTEL

Discrete Probability Distributions

Histogram Example of the Binomial Distribution

Symmetric when $p = 0.5$

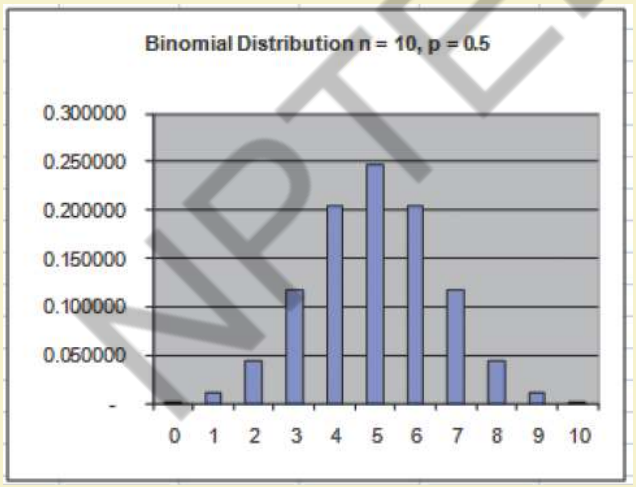


Figure 5.11a

NPTEL

Discrete Probability Distributions

Histogram Examples of the Binomial Distribution

Positively skewed when $p < 0.5$

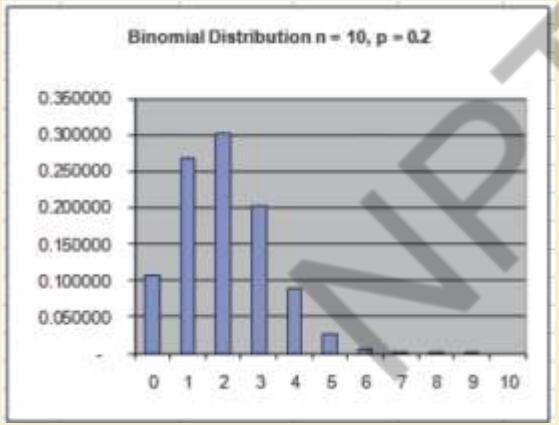


Figure 5.11b

Negatively skewed when $p < 0.5$

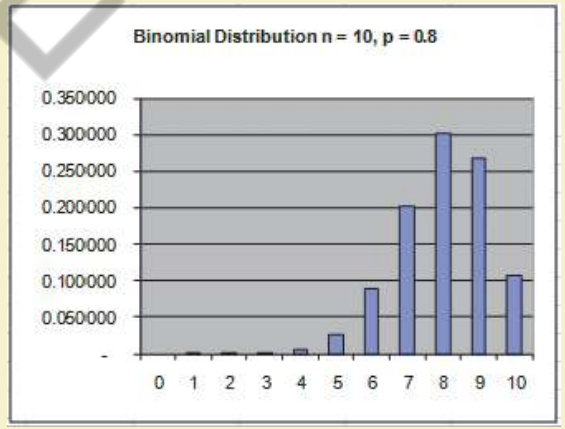


Figure 5.11c

NPTEL

Discrete Probability Distributions

Poisson Distribution

- Models the number of occurrences in some unit of measure (often time or distance).
- There is no limit on the number of occurrences.
- The average number of occurrence per unit is a constant denoted as λ.

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!}, & \text{for } x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

NPTEL

Discrete Probability Distributions

Example 5.26 Computing Poisson Probabilities

- Suppose the average number of customers arriving at a Subway restaurant during lunch hour is λ =12 per hour.
- The probability that exactly *x* customers arrive during the hour is given by the Poisson distribution.
- Find the probability that exactly 5 arrive during lunch hour:
- $f(5) = e^{-12}(12^5)/5!$

  $= (0.000006144)(248,832)/120$

  $= 0.1274$

NPTEL

Discrete Probability Distributions

Example 5.27 Using Excel's Poisson Distribution Function

POISSON.DIST(*x, mean, cumulative*)

True: *F(x)*

False: *f(x)*

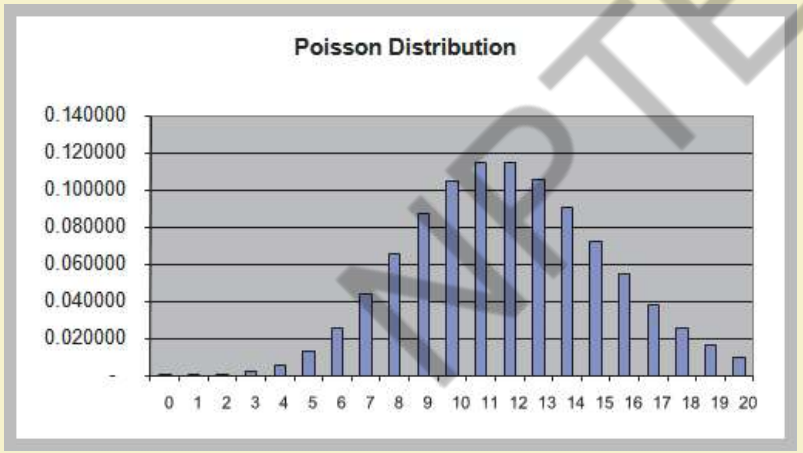


Figure 5.13

	A	B	C
1	Poisson Probabilities		
2			
3	Mean	12	
4			
5	x	f(x)	F(x)
6	0	0.00001	0.00001
7	1	0.00007	0.00008
8	2	0.00044	0.00052
9	3	0.00177	0.00229
10	4	0.00531	0.00760
11	5	0.01274	0.02034
12	6	0.02548	0.04582
13	7	0.04368	0.08950
14	8	0.06552	0.15503
15	9	0.08736	0.24239
16	10	0.10484	0.34723
17	11	0.11437	0.46160
18	12	0.11437	0.57597
19	13	0.10557	0.68154
20	14	0.09049	0.77202
21	15	0.07239	0.84442
22	16	0.05429	0.89871
23	17	0.03832	0.93703
24	18	0.02555	0.96258
25	19	0.01614	0.97872
26	20	0.00968	0.98840

Figure 5.12

NPTEL

Discrete Probability Distributions

Analytics in Practice: Using the Poisson Distribution for Modeling Bids on Priceline

- Pricing strategies for Kimpton hotels on Priceline is modeled using a Poisson distribution.
- The number of bids placed per day 3 days before arrival is $f(x) = e^{-6.3}(6.3^x)/x!$.
- Using the model increased sales 11% in one year.



NPTEL

Continuous Probability Distributions

	A	B
1	Change in DJIA	Subjective Probability
2	-20%	0.01
3	-15%	0.05
4	-10%	0.08
5	-5%	0.15
6	0%	0.2
7	5%	0.25
8	10%	0.18
9	15%	0.06
10	20%	0.02
11		

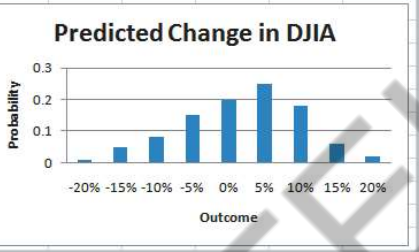


Figure 5.6

Change in DJIA using
5% increments

	A	B
13	Change in DJIA	Subjective Probability
14	-20%	0.0050
15	-17.5%	0.0159
16	-15%	0.0250
17	-12.5%	0.0334
18	-10%	0.0400
19	-7.5%	0.0584
20	-5%	0.0750
21	-2.5%	0.0884
22	0%	0.1000
23	2.5%	0.1134
24	5%	0.1250
25	7.5%	0.1084
26	10%	0.0900
27	12.5%	0.0609
28	15%	0.0300
29	17.5%	0.0209
30	20%	0.0100

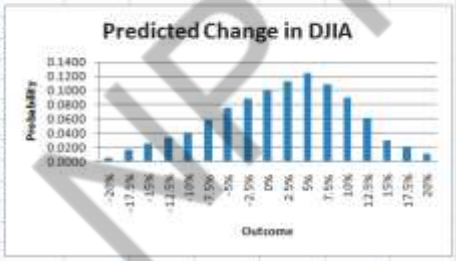


Figure 5.14

Change in DJIA using 2.5% increments

Approaching a smooth curve

NPTEL

Continuous Probability Distributions

Probability density function

- A curve described by a mathematical function that characterizes a continuous random variable

Properties of a probability density function

- $f(x) \geq 0$ for all values of x
- Total area under the density function equals 1.
- $P(X = x) = 0$
- Probabilities are only defined over an interval.
- $P(a \leq X \leq b)$ is the area under the density function between a and b.

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

NPTEL

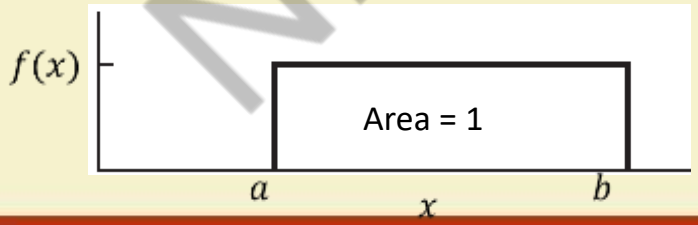
Continuous Probability Distributions

Uniform Distribution

- All outcomes between a minimum (a) and a maximum (b) are equally likely.

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{x-a}{b-a}, & \text{if } a \le x \le b \\ 1, & \text{if } b < x \end{cases}$$



NPTEL

Continuous Probability Distributions

Uniform Distribution

- Expected Value = $\mathrm{EV}[X] = \frac{(a+b)}{2}$

- Variance = $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$

NPTEL

Continuous Probability Distributions

Example 5.28

Computing Uniform Probabilities

- Sales revenue for a product varies uniformly each week between $1000 and $2000.
- $f(x)$ = 1/(2000-1000)

  = 1/1000

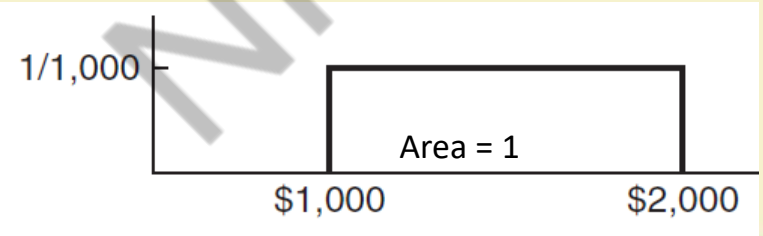


Figure 5.15

NPTEL

Continuous Probability Distributions

Example 5.28 (continued)

Computing Uniform Probabilities

- Find the probability sales revenue will be less than $1,300.
- $P(X < 1300) = (1300\text{-}1000)(1/1000) = 0.30$

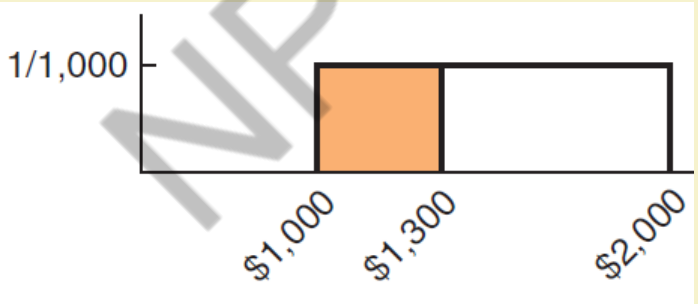


Figure 5.16

NPTEL

Continuous Probability Distributions

Example 5.28 (continued) Uniform Probabilities

- Find the probability that revenue will be between $1,500 and $1,700.



Figure 5.17

- $P(1500 \leq X \leq 1700) = P(X \leq 1700) - P(X \leq 1500)$

$= F(1700) - F(1500)$

$= 300/1000 - 500/1000$

$=0.20$

NPTEL

Continuous Probability Distributions

Normal Distribution

- f(x) is a bell-shaped curve
- Characterized by 2 parameters
  μ (mean)
  σ (standard deviation)
- Properties
  1. Symmetric
  2. Mean = Median = Mode
  3. Unbounded
  4. Empirical rules apply

NPTEL

Continuous Probability Distributions

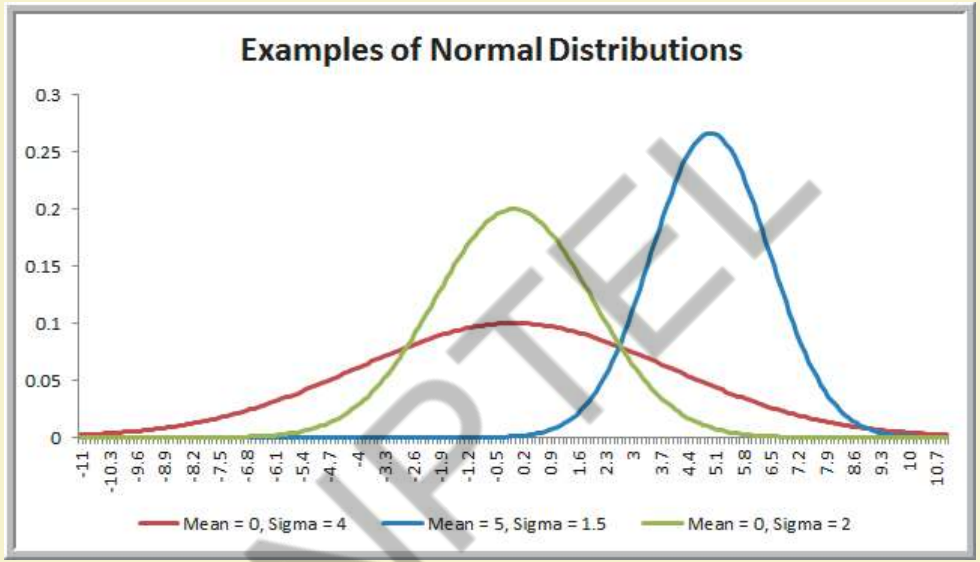


Figure 5.18

=NORM.DIST(*x, mean, stdev, cumulative*)

True: *F(x)*

False: *f(x)*

NPTEL

Continuous Probability Distributions

Example 5.29 Using NORM.DIST to Compute Normal Probabilities

- The distribution for customer demand (units per month) is normal with:

  mean = 750

  stdev. = 100

- Find the probability that demand will be:

  a) at most 900 units/month

  b) exceed 700 units/month

  c) be between 700 and 900 units/month

NPTEL

Continuous Probability Distributions

Example 5.29 (continued) Using NORM.DIST to Compute Normal Probabilities

Cumulative probabilities are computed as:

=NORM.DIST(*x*, 750, 100, true)

a) $P(X < 900) = 0.9332$

b) $P(X > 700) = 1 - 0.3085 = 0.6915$

c) $P(700 < X < 900) = 0.9332 - 0.3085 = 0.6247$

	A	B
1	Normal Probabilities	
2		
3	Mean	750
4	Standard Deviation	100
5		
6	x	F(x)
7	500	0.0062
8	550	0.0228
9	600	0.0668
10	650	0.1587
11	700	0.3085
12	750	0.5000
13	800	0.6915
14	850	0.8413
15	900	0.9332
16	950	0.9772
17	1000	0.9938

Figure 5.19

NPTEL

Continuous Probability Distributions

Example 5.29 (continued) Using NORM.DIST to Compute Normal Probabilities

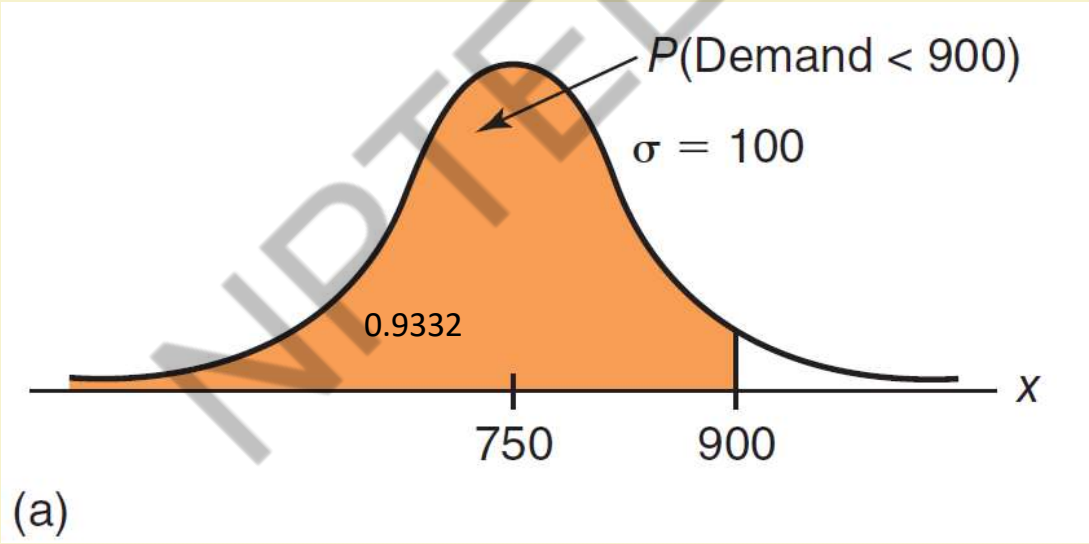


Figure 5.20a

NPTEL

Continuous Probability Distributions

Example 5.29 (continued) Using NORM.DIST to Compute Normal Probabilities

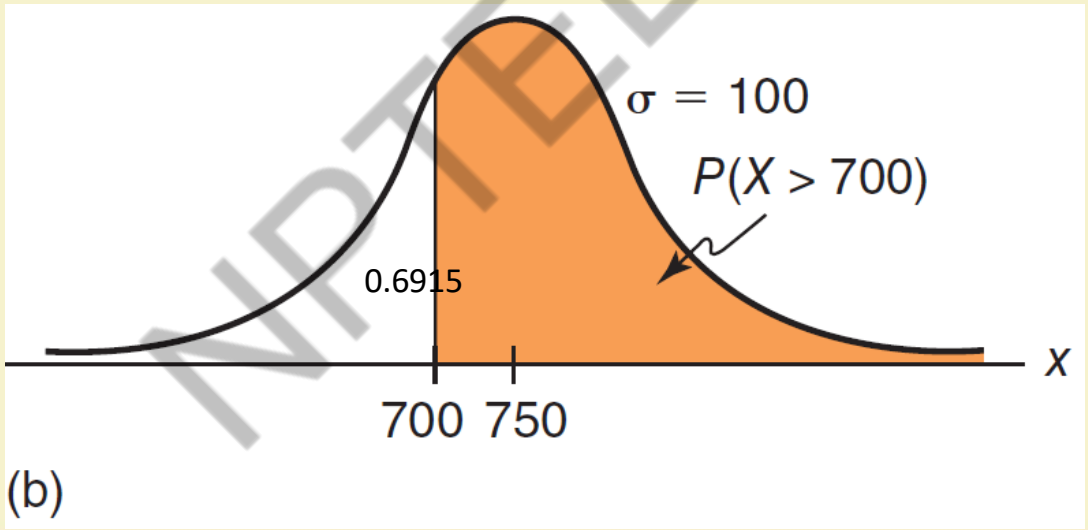


Figure 5.20b

NPTEL

Continuous Probability Distributions

Example 5.29 (continued) Using NORM.DIST to Compute Normal Probabilities

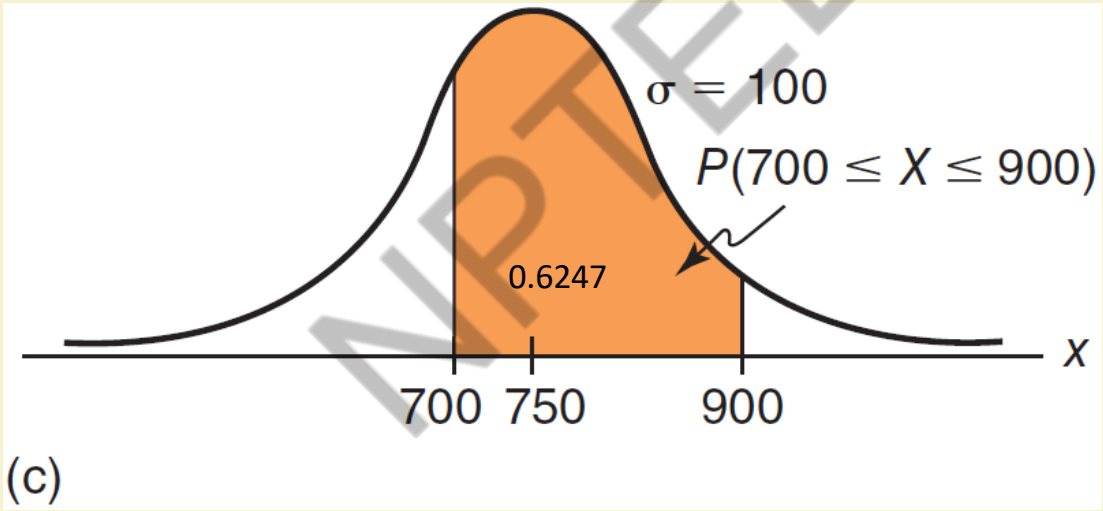


Figure 5.20c

NPTEL

Continuous Probability Distributions

Example 5.30 Using the NORM.INV Function

=NORM.INV(*probability, mean, stdev*)
provides the *x* value with *F(x)= probability*

- What level of demand would be exceeded at most 10% of the time?
- Find *x* such that *F(x)* = 90%

= NORM.INV(0.90, 750, 100)

results in *x* = 878.155

NPTEL

Continuous Probability Distributions

Example 5.30 (continued)

Using the NORM.INV Function

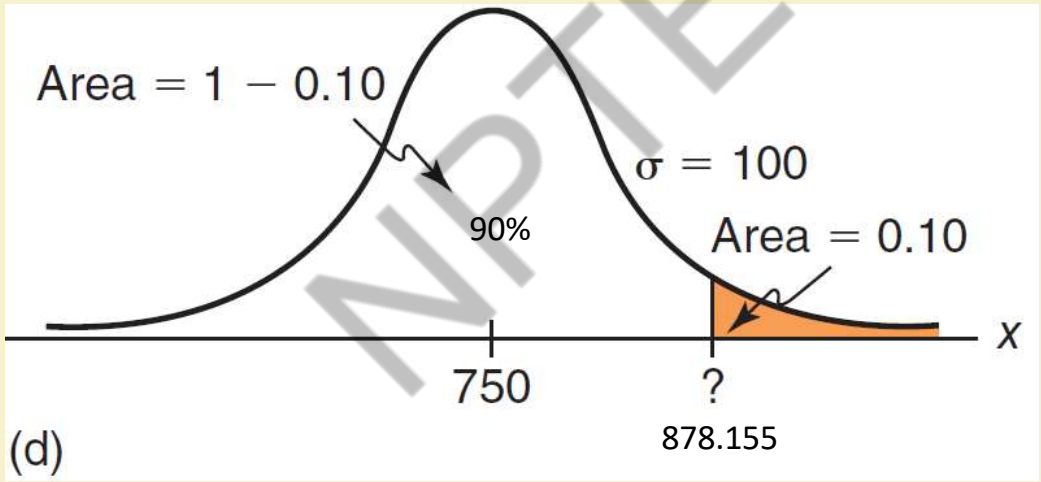


Figure 5.20d

NPTEL

Continuous Probability Distributions

Standard Normal Distribution

Z is the standard normal random variable with:

Mean = 0

Stdev = 1

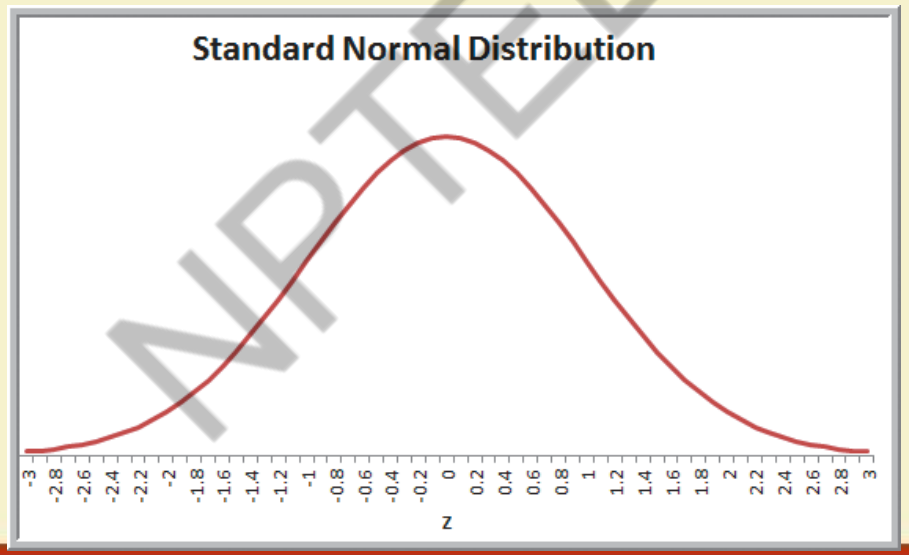


Figure 5.21

NPTEL

Continuous Probability Distributions

Example 5.31 Computing Probabilities with the Standard Normal Distribution

- Verify the empirical rules using Excel.
- $P(-1 < Z < 1)$

  = NORMS.DIST(1) – NORMS.DIST(-1)

  = 0.84134 – 0.15866

  = 0.6827

  ~ 68%

	A	B	C	D	E	F	G	H
1	Standard Normal Probabilities							
2								
3	z	F(z)		a	b	F(a)	F(b)	F(b) - F(a)
4	-3	0.00135		-1	1	0.15866	0.84134	0.6827
5	-2	0.02275		-2	2	0.02275	0.97725	0.9545
6	-1	0.15866		-3	3	0.00135	0.99865	0.9973
7	0	0.50000						
8	1	0.84134						
9	2	0.97725						
10	3	0.99865						

Figure 5.22

NPTEL

Continuous Probability Distributions

Example 5.32 Computing Probabilities with the Standard Normal Tables

- From Example 5.29, what is the probability that demand will be at least 900 units/month?
- Use the equation: $z = \frac{(x - \mu)}{\sigma}$
- $Z = (900 - 750)/100$
  $= 1.50$
- Using Table 1 in Appendix B, we find:

$P(X < 900) = P(Z < 1.50)$
$= 0.93319$

NPTEL

Continuous Probability Distributions

Exponential Distribution

- Models the time between randomly occurring events (arrivals, machine failures, etc.)

$$f(x) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0$$

$$F(x) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0$$

$$\mu = \frac{1}{\lambda}$$

- where λ is the mean rate of occurrences (from the discrete Poisson distribution)

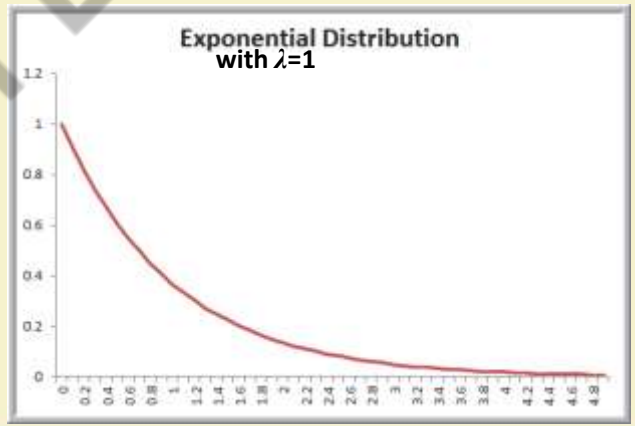


Figure 5.23

NPTEL

Continuous Probability Distributions

Example 5.33 Using the Exponential Distribution

- The mean time to failure of a critical engine component is μ = 8,000 hours.

What is the probability of failing before 5000 hours?

- *P(X < x)* =EXPON.DIST(*x, lambda, cumulative*)
- Since $\mu = \frac{1}{\lambda}$, we can solve for $\lambda = \frac{1}{\mu}$

λ = 1/8000

P(x < 5000) =EXPON.DIST(5000, 1/8000, true)

= 0.4647

NPTEL

Continuous Probability Distributions

Example 5.33 (continued)

Using the Exponential Distribution

	A	B
1	**Exponential Probabilities**	
2		
3	**Mean**	8000
4	**Lambda**	0.0001
5		
6	**x**	**F(x)**
7	0	0
8	1000	0.1175
9	2000	0.2212
10	3000	0.3127
11	4000	0.3935
12	5000	0.4647
13	6000	0.5276
14	7000	0.5831
15	8000	0.6321
16	9000	0.6753
17	10000	0.7135
18	11000	0.7472

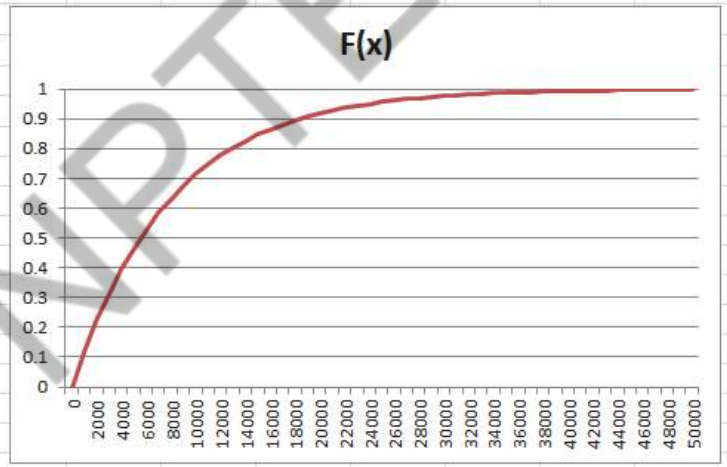


Figure 5.24

NPTEL

Continuous Probability Distributions

Other Useful Continuous Distributions

- Triangular Distribution
- Lognormal Distribution
- Beta Distribution

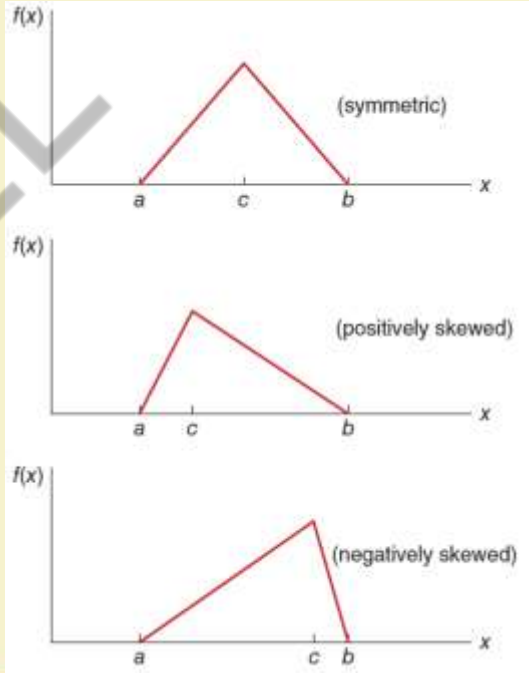


Figure 5.25

NPTEL

Random Sampling from Probability Distributions

- Many application in Business Analytics require random samples from specific probability distributions.
- For example, cumulative discounted cash flow can be described using probability distributions for sales, sales growth rate, operating expenses, and inflation factors (all uncertain variables).
- Random numbers are the basis for generating random samples from probability distributions.
- Random numbers are uniformly distributed on the interval from 0 to 1.

NPTEL

Random Sampling from Probability Distributions

Example 5.34 Sampling from the Distribution of Dice Outcomes

Probability distribution

x	f(x)	F(x)
2	0.028	0.028
3	0.056	0.083
4	0.083	0.167
5	0.111	0.278
6	0.139	0.417
7	0.167	0.583
8	0.139	0.722
9	0.111	0.833
10	0.083	0.917
11	0.056	0.972
12	0.028	1.000

Intervals for random sampling

Interval	Outcome
0 to 0.028	2
0.028 to 0.083	3
0.083 to 0.167	4
0.167 to 0.278	5
0.278 to 0.417	6
0.417 to 0.583	7
0.583 to 0.722	8
0.722 to 0.833	9
0.833 to 0.917	10
0.917 to 0.972	11
0.972 to 1	12

NPTEL

Random Sampling from Probability Distributions

Example 5.34 (continued) Sampling from the Distribution of Dice Outcomes

=RAND( ) generates random numbers in Excel

	A	B
1	Random Numbers	
2		
3	Sample	Random Number
4	1	0.681423018
5	2	0.835253743
6	3	0.438867243
7	4	0.11755569
8	5	0.731253287
9	6	0.584840908
10	7	0.450591681
11	8	0.119527366
12	9	0.778954333
13	10	0.833953932

Outcome = 8 since 0.681 is between 0.583 and 0.722

Outcome = 4 since 0.119 is between 0.083 and 0.167

Figure 5.26

NPTEL

Random Sampling from Probability Distributions

Example 5.35 Using the VLOOKUP Function

- Generate a random sample of Changes in DJIA.
- First compute *F(x)*
- Assign intervals to outcomes
- Generate random numbers using =RAND( )

=VLOOKUP(H2, $E2:$G$10, 3)

	A	B	C	D	E	F	G	H	I
1	Change in DJIA	f(x)	F(x)		Interval		Change in DJIA	Random Number	Outcome
2	-20%	0.01	0.01		0	0.01	-20%	0.681423018	5%
3	-15%	0.05	0.06		0.01	0.06	-15%	0.835253743	10%
4	-10%	0.08	0.14		0.06	0.14	-10%	0.438867243	0%
5	-5%	0.15	0.29		0.14	0.29	-5%	0.11755569	-10%
6	0%	0.2	0.49		0.29	0.49	0%	0.731253287	5%
7	5%	0.25	0.74		0.49	0.74	5%	0.584840908	5%
8	10%	0.18	0.92		0.74	0.92	10%	0.450591681	0%
9	15%	0.06	0.98		0.92	0.98	15%	0.119527366	-10%
10	20%	0.02	1		0.98	1	20%	0.778954333	10%
11								0.833953932	10%

Figure 5.27

NPTEL

Random Sampling from Probability Distributions

Example 5.36

Using Excel's Random Number Generation Tool

Generate 100 outcomes from a Poisson distribution with a mean of 12.

Data

Data Analysis

Random Number Generation

Number of Variables: 1

Number of Random Numbers: 100

Distribution: Poisson

Parameter: Lambda = 12

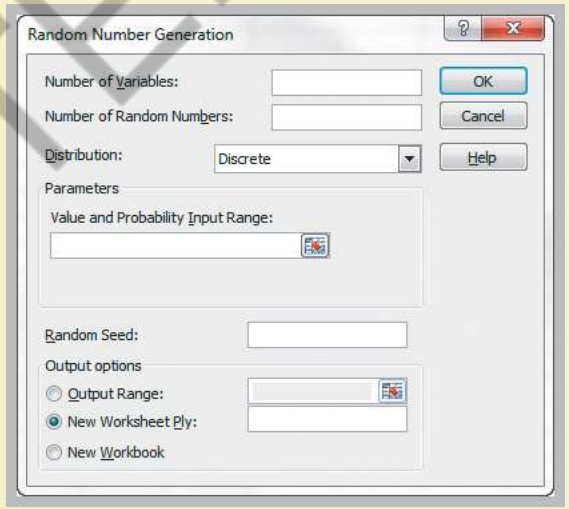


Figure 5.28

NPTEL

Random Sampling from Probability Distributions

Example 5.36 (continued) Using Excel's Random Number Generation Tool

Histogram of 100 random outcomes

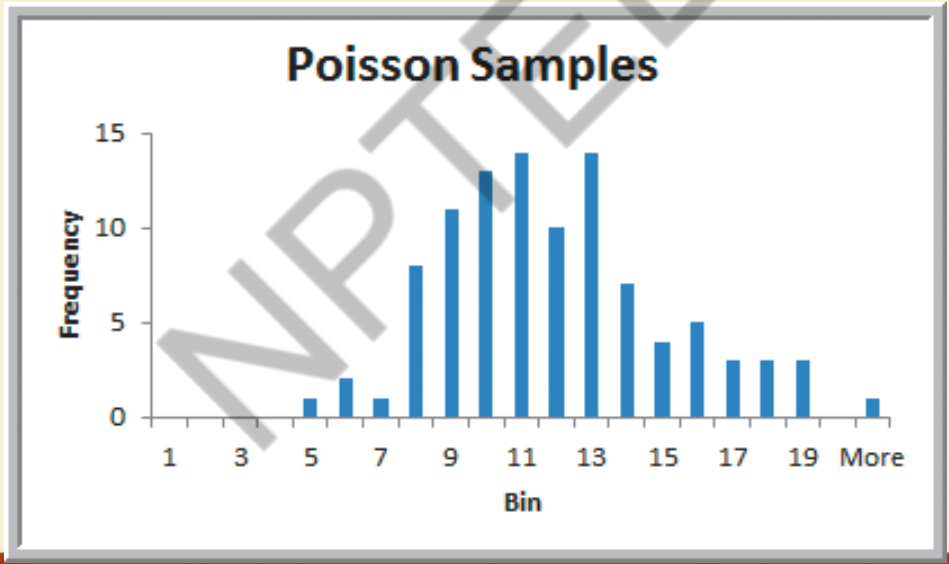


Figure 5.29

NPTEL

Random Sampling from Probability Distributions

Probability Distributions available in Excel's Data Analysis Toolpak under *Random Number Generation*

- Bernoulli
- Binomial
- Discrete
- Normal
- Patterned
- Poisson
- Uniform

NPTEL

Random Sampling from Probability Distributions

Sampling From Various Probability Distributions Using Standard Excel Functions

Uniform: =RANDBETWEEN*(a, b)*

You can also replace *probability* with *RAND( )* in many inverse probability distribution functions:

Binomial: =BINOM.INV*(trials, p, RAND( ))*

Normal: =NORM.INV*(RAND( ), mean, stdev)*

Logormal: =LOGNORM.INV*(RAND( ), mean, stdev)*

Beta: =BETA.INV*(RAND( ), alpha, beta, A, B)*

NPTEL

Random Sampling from Probability Distributions

Example 5.37 A Sampling Experiment for Evaluating Capital Budgeting Projects

- In finance, one way of evaluating capital budgeting projects is to compute:
- Profitability index $PI = PV / I$

where PV is the present value of future cash flows

and I is the initial investment

NPTEL

Random Sampling from Probability Distributions

Example 5.37 (continued) A Sampling Experiment for Evaluating Capital Budgeting Projects

- Suppose *PV* is normally distributed with a mean of $12 million and standard deviation of $2.5 million.
- *I* is also normally distributed with a mean of $3 million and standard deviation of $0.8 million.
- Use a sampling experiment to generate a probability distribution for *PI = PV / I*.

NPTEL

Random Sampling from Probability Distributions

Example 5.37 (continued)

- For *PV:* =NORM.INV(RAND( ), 12, 2.5)
- For *I*: =NORM.INV(RAND( ), 3, 0.8)

	A	B	C	D	E
1	Profitability Index Analysis				
2					
3		Mean	Standard Deviation		
4	PV	12	2.5		
5	I	3	0.8		
6					
7	Experiment	PV	I	PI	Mean
8	1	8.396743042	3.573822001	2.349513501	4.762285
9	2	11.7446542	3.66554571	3.204067043	
10	3	11.76586862	3.554538257	3.310097619	
11	4	11.44456518	3.33708406	3.429510606	
12	5	9.373641185	3.692222659	2.538752955	
13	6	10.47906344	2.598868941	4.0321631	
14	7	14.31716958	3.203954788	4.46859289	
15	8	8.901052248	0.729081227	12.20858791	
16	9	13.99414343	3.180751244	4.399634662	
17	10	12.5758327	3.513579887	3.579207847	

PI = PV/I

Profitability Index mean = 4.76

Figure 5.30

NPTEL

Random Sampling from Probability Distributions

Example 5.37 (continued) A Sampling Experiment for Evaluating Capital Budgeting Projects

- Profitability Index is skewed to the right

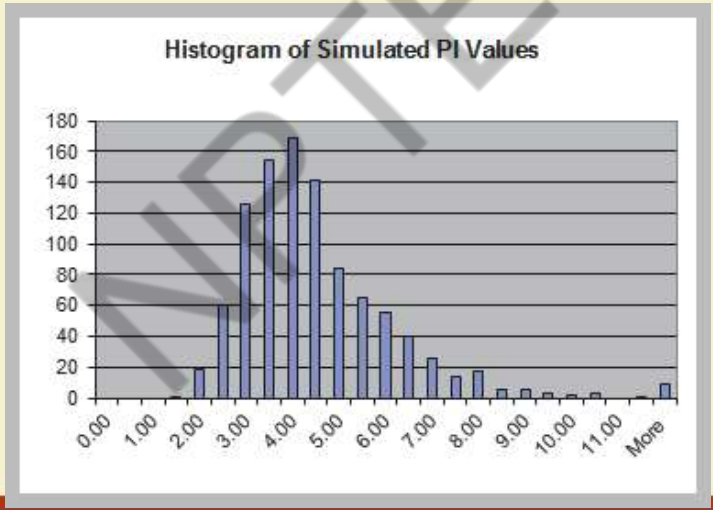


Figure 5.31

NPTEL

Random Sampling from Probability Distributions

Risk Solver Platform

Probability Distribution Functions

Distribution	RSP Function
Bernoulli	PsiBernoulli(*probability*)
Binomial	PsiBinomial(*trials, probability*)
Poisson	PsiPosson(*mean*)
Uniform	PsiUniform(*lower, upper*)
Normal	PsiNormal(*mean, standard deviation*)
Exponential	PsiExponential(*mean*)
Discrete Uniform	PsiDisUniform(*values*)
Geometric	PsiGeometric(*probability*)
Negative Binomial	PsiNegBinomial(*successes, probability*)
Hypergeometric	PsiHyperGeo(*trials, success, population size*)
Triangular	PsiTriangular(*minimum, most likely, maximum*)
Lognormal	PsiLognormal(*mean, standard deviation*)
Beta	PsiBeta(*alpha, beta*)

Table 5.1

NPTEL

Random Sampling from Probability Distributions

Example 5.38 Using *Risk Solver Platform* Distribution Functions

- An energy company is considering offering a new product and needs to estimate the growth in PC ownership. The expected growth rates are:
- Minimum = 5%
- Most likely = 7.7%
- Maximum =10%
- Generate 500 samples of PC ownership growth rate using: =PsiTriangular(5%, 7.7%, 10%)

NPTEL

Random Sampling from Probability Distributions

Example 5.38 (continued) Using *Risk Solver Platform* Distribution Functions

	A	B
1	PC Ownership Growth Rates	
2		
3	Minimum	5%
4	Most Likely	7.70%
5	Maximum	10%
6		
7	Sample	Growth Rate
8	1	6.37%
9	2	8.77%
10	3	7.01%
11	4	7.36%
12	5	8.21%
13	6	8.01%
14	7	7.39%
15	8	7.29%
16	9	7.68%
17	10	8.97%
18	11	7.21%
19	12	8.31%
20	13	7.60%
21	14	6.62%
504	497	8.09%
505	498	8.49%
506	499	8.03%
507	500	8.20%

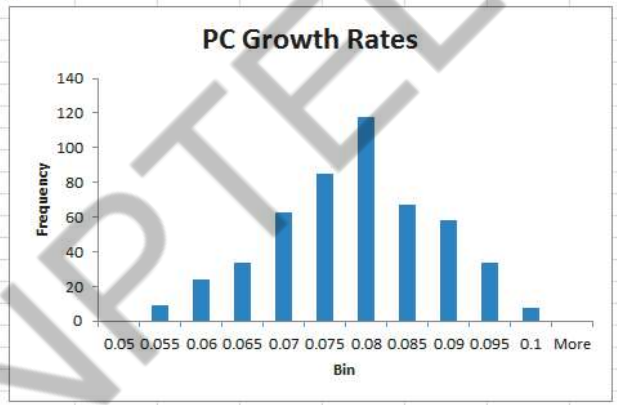


Figure 5.32

NPTEL

Data Modeling and Distribution Fitting

Example 5.39 Analyzing Airline Passenger Data

- Sample data on passenger demand for 25 flights

Passengers	Bin	Frequency
36	30	0
55	30	0
47	32.5	0
45	32.5	0
48	35	1
43	37.5	0
42	40	2
56	42.5	2
40	45	6
47	47.5	3
44	50	5
46	52.5	3
53	55	2
45	57.5	1
44	60	0
45	More	0
45		
41		
47		
46		
40		
46		
42		
41		
58		

Passengers	
Mean	45.68
Standard Error	1.043584
Median	45
Mode	45
Standard Deviation	5.217918
Sample Variance	27.22667
Kurtosis	0.707219
Skewness	0.823163
Range	22
Minimum	36
Maximum	58
Sum	1142
Count	25



Figure 5.33

Can we assume normally distributed?

NPTEL

Data Modeling and Distribution Fitting

Example 5.40 Analyzing Airport Service Times

- Sample data on service times for 812 passengers at an airport's ticketing counter

	A	B	C	D
1	Service Times at an Airport Ticketing Counter			
2				
3	Times (sec.)		*Times (sec.)*	
4	227			
5	83		Mean	126.2783
6	10		Standard Error	3.691221
7	158		Median	88
8	350		Mode	83
9	15		Standard Deviation	105.1836
10	63		Sample Variance	11063.59
11	224		Kurtosis	8.707526
12	96		Skewness	2.413577
13	61		Range	867
14	61		Minimum	9
15	65		Maximum	876
16	91		Sum	102538
17	133		Count	812

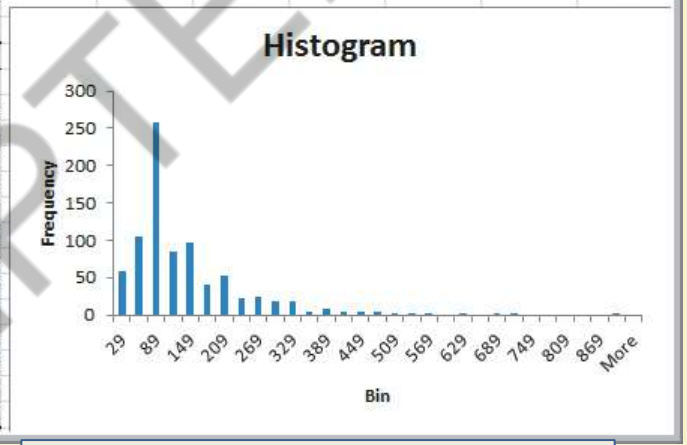


Figure 5.34

Can we assume normally distributed?

NPTEL

Data Modeling and Distribution Fitting

Goodness of Fit

- The basis for fitting data to a probability distribution
- Attempts to draw conclusions about the *nature* of the distribution
- Three statistics measure goodness of fit:

  Chi-square

  Kolmogorov-Smirnov

  Anderson-Darling

NPTEL

Data Modeling and Distribution Fitting

Example 5.41

Fitting a Distribution to Airport Service Times

1. Highlight the data

   Risk Solver
   Tools
   Fit

2. *Fit Options* dialog

   Type: Continuous
   Test: Kolmorgov-Smirnov

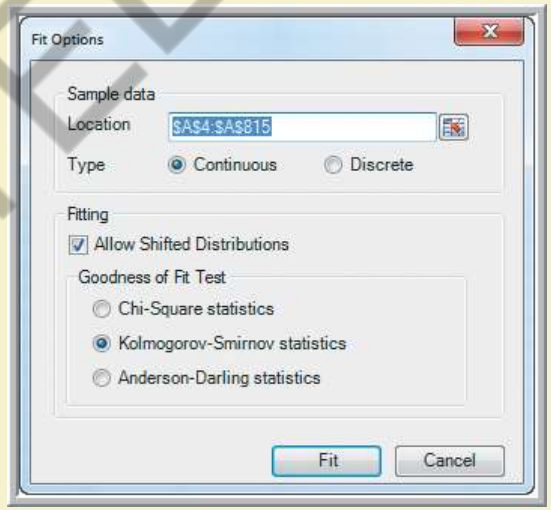


Figure 5.35

NPTEL

Data Modeling and Distribution Fitting

Example 5.41 (continued)

Fitting a Distribution to Airport Service Times

- Erlang is the best-fitting distribution

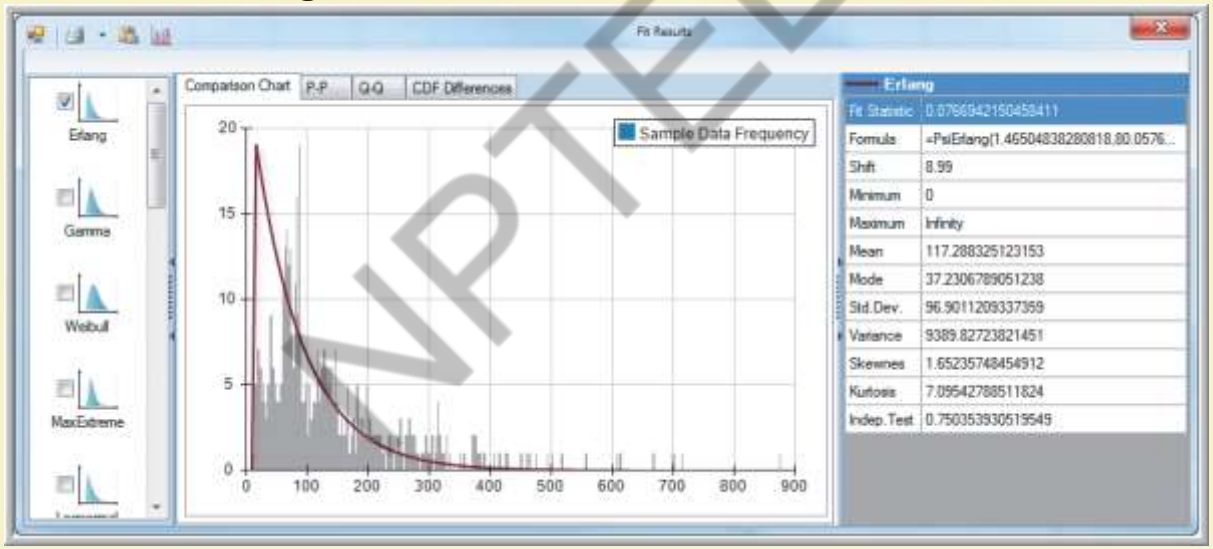


Figure 5.36

NPTEL

Data Modeling and Distribution Fitting



Analytics in Practice: The Value of Good Data Modeling in Advertising

- Gross's model:
- A mathematical model that relates the relative contributions of creative and media dollars to total advertising effectiveness.
- Often used to identify the best number of ads to purchase.
- Analysis found that the optimal number of ads can vary significantly depending on the shape of the distribution of effectiveness for a single ad.

NPTEL

Chapter 5 – Key Terms

- Bernoulli distribution
- Binomial distribution
- Complement
- Conditional probability
- Continuous random variable
- Cumulative distribution function
- Discrete random variable
- Discrete uniform distribution
- Empirical probability distribution
- Event
- Expected value
- Experiment
- Exponential distribution
- Goodness of fit
- Independent events

NPTEL

Chapter 5 – Key Terms (continued)

- Multiplication law of probability
- Mutually exclusive
- Normal distribution
- Outcome
- Poisson distribution
- Probability
- Probability density function
- Probability distribution
- Probability mass function
- Random number
- Random number seed
- Random variable
- Random variate
- Sample space
- Standard normal distribution
- Uniform distribution
- Union

NPTEL

Case Study
Performance Lawn Equipment (5)





- Recall that PLE produces lawnmowers and a medium size diesel power lawn tractor.
- Determine probability distributions appropriate for modeling mower failures and blade weights.
- Determine the sampling distribution of the mean mower failures for a sample size of 100.
- Determine the sampling distribution of the proportion blade weights below specification for a sample size of 350.
- Analyze your results and write up a formal report.

NPTEL