Computer Applications

Probability

Year 2

Semester 1

Dr. Jane M. Horgan, School of Computer Applications.

Web: http://www.compapp.dcu.ie/~jhorgan

Contents

1 WHAT IS STATISTICS? 9

2 BASIC CONCEPTS OF PROBABILITY 18

3 AXIOMS OF PROBABILITY 29

4 CONDITIONAL PROBABILITY 37

5 RELIABILITY 42

6 BAYES THEOREM 45

7 RANDOM VARIABLES 51

8 GEOMETRIC DISTRIBUTION 56

9 BINOMIAL DISTRIBUTION 66

10 SAMPLING WITHOUT REPLACEMENT 71

11 EXPECTATIONS 81

12 VARIANCES 86

13 PROPERTIES OF EXPECTATIONS 92

14 POISSON DISTRIBUTION 95

15 APPLICATIONS OF THE POISSON 98

16 SAMPLING INSPECTION 104

17 DOUBLE SAMPLING SCHEMES 114

18 CONTINUOUS DISTRIBUTIONS 118

19 EXPECTATION OF CONTINUOUS VARIABLES127

20 THE NORMAL DISTRIBUTION 130

21 APPLICATIONS OF NORMAL DISTRIBUTION 133

22 INEQUALITIES 136

STATISTICS YEAR 2 SEMESTER 1

Course Content

1. Descriptive Statistics: Populations and samples. Measures of central tendency and dispersion.

2. Introduction to Probability Theory: Sample space. Some probability laws. Conditional probability. Bayes' theorem.

3. Random Variables and Expectation: Discrete and continuous random variables. Laws of expectation. Markov's and Chebychev's inequalities. Moment generating function. Jointly distributed random variables.

4. Discrete Distributions: Bernouilli, hypergeometric, binomial, geometric and Poisson distributions. Limiting distributions of hypergeometric and binomial. Markov property.

5. Reliability.

6. Sampling Inspection: Single sampling schemes. Double sampling schemes. Operating Characteristic. Average outgoing quality. Consumer's and producer's risks.

7. Continuous Distribution: Exponential and normal. Areas under the normal curve. Normal approximation to the binomial.

Reading List

Main Texts:

- Horgan Jane M. *Probability with R: An Introduction with Computer Science Applications*, John Wiley and Sons.
- Venables, W. N. Smith, D.M., and the R Development, *An Introduction to R: A Programming Environment for Data Analysis and Graphics*, http://cran-project.org/manuals

Supplementary Reading:

Allen, Arnold, O. *Probability, Statistics, and Queueing Theory with Computer Science Applications*, Academic Press. 1978

Berenson, M. L. and Levine, D. M. *Basic Business Statistics.* Prentice Hall, 1996.

Crawley, M. *Statistics an Introduction using R*, 2007

Chatfield, C. *Statistics for Technology*, Chapman and Hall, 1983.

Dalgaard, Peter (2002) *Introductory Statistics with R*, Springer

Dougherty, E. R. *Probability and Statistics for Engineering, Computing and Physical Sciences*, Prentice-Hall, 1990.

Freund, J. E. *Mathematical Statistics*, Prentice-Hall, 1980.

Greer, A. *Statistics for Engineers*, Prentice-Hall.

Kaplan, M. and Kaplan, E. *Adventures in Probability* Viking, 2006

Kinney, J. J. *Probability: An Introduction with Statistical Applications*, Wiley, 1997.

MacDonald J, and Braun, J.*Data Analysis and Graphics using R*, Cambridge, 2006.

Montgomery, D. C. *Applied Statistics and Probability for Engineers*, Wiley, 1994.

Olofsson, Peter *The Little Numbers that Rule Our Lives.* 2007

Reilly James *Understanding Statistics and its Applications in Business, Science and Engineering* Folans, 1997

Trevidi, Kishor S. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications.* Prentice-Hall. 1982.

Wetherill G. Barrie. *Elementary Statistical Methods*, Chapman and Hall, 1981.

White, J., Yeats, A. and Skipworth, G. *Tables for Statisticians*

The subject of Statistics provides answers to questions such as:

1. What data need be collected?
2. How can resources be used efficiently to collect the data?
3. How can the data be presented so as to convey their salient features?
4. What conclusions can be drawn from these data and what is the degree of certainty of these conclusions?
5. What actions should be taken on the basis of the conclusions drawn from these data?

Applications of Probability and Statistics:

1. Computing: Evaluating algorithms; evaluating queueing systems for computer usage; cpu time; response time; bytecode usage by different Java compilers; extracting information from large data sets.
2. Engineering: Improving product design and testing product performance; estimating reliability of components and systems.
3. Quality Control: Evaluating quality through sampling; process control.
4. Actuarial Science: Determining premium rates; designing pension plans.
5. Business, Accounting and Industry. Estimating sales; auditing.
6. Economics; Measuring indicators such as trade, size of labour force, and standard of living; long and short range forecasts of economic indicators.
7. Health and Medicine: Drug testing.
8. Psychology: IQ testing.
9. Opinion Polls: Who will win the next general election?

1 WHAT IS STATISTICS?

STATISTICS deals with numerical data, in particular the:

- COLLECTION
- PRESENTATION
- DESCRIPTION
- INTERPRETATION

Figure 1: Boxplots

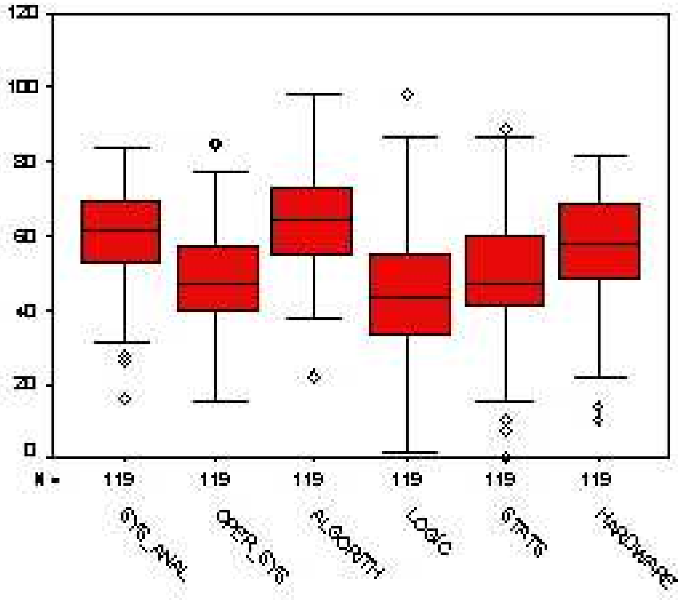
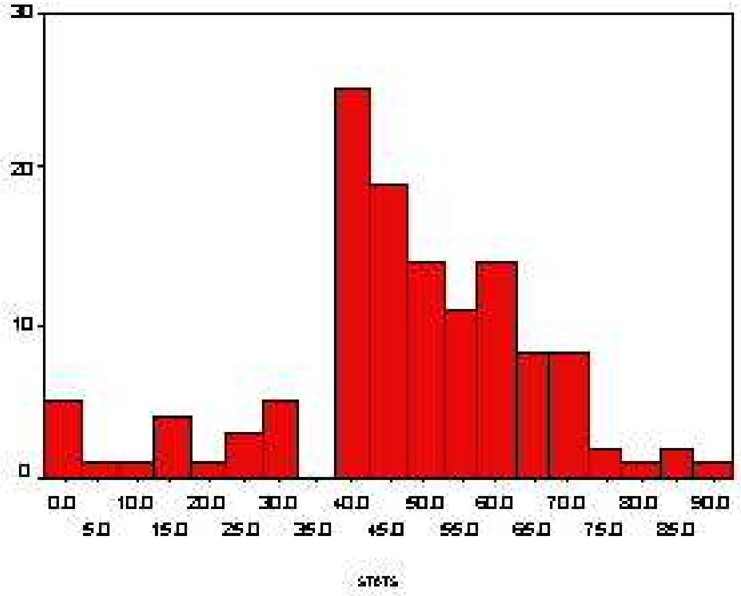


Table 1: Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
STATS	125	89.00	.00	89.00	47.3920	18.0147
SYS ANAL	121	67.00	16.00	83.00	59.2645	12.6805
OPER SYS	121	70.00	15.00	85.00	48.7190	14.0465
ALGORITH	121	76.00	22.00	98.00	64.6198	14.2128
LOGIC	123	96.00	2.00	98.00	44.9675	18.4297
HARDWARE	120	71.00	10.00	81.00	56.9167	15.1960

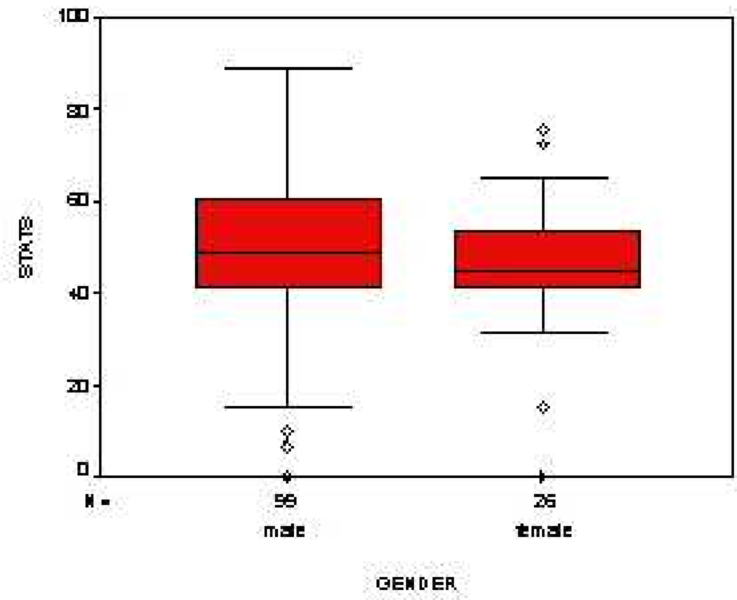
Figure 2: Histgram and Stem−and−Leaf of Statistics Exam



Stem and Leaf Diagram

7.00	Extremes		(=< 10)
.00	1	.	
4.00	1	.	5567
2.00	2	.	23
4.00	2	.	6689
3.00	3	.	022
.00	3	.	
34.00	4	.	0000000011111111111222222333344444
16.00	4	.	5555666677889999
13.00	5	.	0001112233444
10.00	5	.	5567778899
12.00	6	.	000111222233
9.00	6	.	557777999
5.00	7	.	00022
3.00	7	.	568
1.00	8	.	4
1.00	8	.	7
1.00	Extremes		(>= 89)

Stem width: 10.00 Each leaf: 1 case(s)


100
80
60
40
20
0
STATS
N =
59
26
male
female
GENDER

Example:

A manufacturer of minicomputer systems is interested in improving its customer support services. As a first step its marketing department has been charged with the responsibility of summarizing the extent of customer problems in terms of system downtime. The twenty three most recent customers were surveyed and the amount of downtime (in minutes) they had experienced during the previous month was summarised in the following stem and leaf display.

Freq	Stem	Leaf
3	0	012
4	1	2248
7	2	1134599
5	3	00036
3	4	457
1	5	1

1. Re-create all the numbers in the data set.
2. Calculate the quartiles of the data.
3. Construct a box-plot.

DATA PRESENTATION AND DESCRIPTION

- **Summary Statistics:**

  1. *Measures of Central Tendency:* These measure typical or central points in the data. They include:
     (a) *Mean:* Sum of all values divided by the number of cases.
     (b) *Median:* Middle value. 50% of the data lies below this value and 50% above.
     (c) *Mode:* Most commonly occurring value or equivalently, the value with the highest frequency.
  2. *Measures of Dispersion:* These measure the spread or variation in the data. Most commonly used measures include:
     (a) *Standard Deviation (SD):* The square root of the average squared deviations from the mean. This measures how the data values differ from the mean. A small standard deviation implies most values are near the average. A large standard deviation indicates that values are widely spread above and below the average.
     (b) *Range:* Lowest and highest values.
     (c) *Percentiles:* Values that divide cases below which certain percentages of values fall. The 50th percentile is the median.
     (d) *Interquartile Range:* 25th to 75th percentiles, or equivalently the middle 50% of the data.
     (e) *Standard Error (SE):* Standard deviation of the mean.

- **Graphical Illustrations:**

  1. *Boxplot:*
     This is a graphical summary based on the median, quartiles and extreme values. Often called the *Box and Whiskers Plot*, the box represents the interquartile range which contains 50% of cases. The whiskers are lines that extend from the box to the highest and lowest values. A line across the box indicates the median. Extreme values are cases more than 1.5 box lengths from the upper or lower end of the box. The extreme cases are listed on the plot.
  2. **Stem and Leaf**
     This is a depiction of the shap of the data based on the actual numbers observed. Then stem usually depict the 10s and the leaves depict the units.

Basic Terminology

- **Population:**

  The complete set of elements to be covered.

- **Sample:**

  A subset of the population.

- **Parameter:**

  A characteristic of the entire population.

- **Statistic:**

  A characteristic of the sample.

WHAT IS PROBABILITY?

- **Chance.**

- **16th and 17th Century: Games of Chance.**

- **Toss a ‘fair’ die.**

- **Pull a card from a ‘well-shuffled deck’.**

2 BASIC CONCEPTS OF PROBABILITY

- **Experiment** is a process of observation that leads to a single outcome that cannot be predicted with certainty.

  Examples:

  1. Pull a card from a deck
  2. Toss a coin
  3. Response time.

- **Sample Space:** All outcomes of an experiment.
  Usually denoted by **S**.

- **Event** denoted by **E** is any subset of **S**

  1. E = Spades
  2. E = Head
  3. E = Component is functioning

- **P(E)** denotes the probability of the event **E**.

  1. P(E) = P(Spades)
  2. P(E) = P(Head)
  3. P(E) = P(Component is functioning)

Calculating Probabilities

- **CLASSICAL APPROACH**:

Assumes all outcomes of the experiment are equally likely:

$$P(E) = \frac{\text{number of favourable cases}}{\text{number of possible cases}}$$

Example: Roll a fair die.

E = even number

$$P(E) = \frac{3}{6}$$

- **RELATIVE FREQUENCY APPROACH**:

Interprets the probability as the relative frequency of the event over a long series of experiment.

$$P(E) = \frac{\text{no. of times } \mathbf{E} \text{ occurs}}{\text{no. of times experiment is repeated}}$$

Example: Roll a die a large number of times and observe number of times an even number occurs.

$$P(E) = \frac{\text{number of observed evens}}{\text{number of times the die is rolled}}$$

Examples:

1. Toss a coin.
   What is the probability of getting a head?

2. Toss two coins.
   What is the probability of getting at least one head?

3. Select a card from a deck.
   What is the probability that it is a diamond?

4. A group of four integrated-circuit (IC) chips consists of two good chips and two defective chips. If three chips are selected at random from this group, what is the probability that two are defective?
   Solution:
   A natural sample space for this problem consists of all possible three-chip selections from the group of four chips:

   $$S = \{g_1g_2d_1, g_1g_2d_2, g_1d_1d_2, g_2d_1d_2\}.$$

   $E =$ Two of the three selected chips are defective.
   Since the two sample points $g_1d_1d_2$ and $g_2d_1d_2$ are favourable to the event E and since the sample space has four points, we conclude that $P(E) = 2/4 = 1/2$.

Permutations and Combinations:
We have seen that finding $P(E)$ simply involves counting the number of equally likely outcomes favourable to E. However, counting by hand may not be feasible when the sample space is large.

Permutations:
The number of ordered sequences where repetition is not allowed, i.e. no element can appear more than once in the sequence.
Examples:

1. Three elements {1,2,3}.
   How many sequences of two elements from these three?
   (1,2); (1,3); (2,1); (2, 3); (3,1); (3,2).
   Six ordered sequences altogether.

2. Four elements {1,2,3,4}.
   How many sequences of two elements from these four?
   (1,2); (1,3); (1,4) (2,1); (2, 3); (2,4);
   (3,1); (3,2); (3,4); (4,1); (4,2); (4,3).
   Twelve ordered sequences altogether.

3. Four elements {1,2,3,4}.
   How many sequences of three elements from these four?
   (1,2,3); (1,3,2); (1,2,4); (1,4,2); (1,3,4); (1,4,3)
   (2,1,3); (2,3,1); (2,1,4); (2,4,1); (2,3,4); (2,4,3)
   (3,1,2); (3,2,1); (3,2,4); (3,4,2); (3,1,4); (3,4,1)
   (4,1,3); (4,3,1); (4,1,2); (4,2,1); (4,3,2); (4,2,3)
   Twenty-four ordered sequences.

Permutations:

Ordered samples (sequences) of size k from n

$$^{n}P_{k} = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

This is also known as the number of permutations of n distinct objects taken k at a time.
i.e.

$$^{n}P_{k} = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

1. $^{3}P_{2} = 3 * 2 = 6$
2. $^{4}P_{2} = 4 * 3 = 12$
3. $^{4}P_{3} = 4 * 3 * 2 = 24$

Example

Find the probability that a randomly chosen three-letter sequence will not have any repeated letters.

Let $I = \{a, b, \ldots, z\}$ be the alphabet of 26 letters.
Then the sample space is

$$S = \{(\alpha, \beta, \gamma) : \alpha \in I, \beta \in I, \gamma \in I\}$$

and the event of interest is

$$E = \{(\alpha, \beta, \gamma) : \alpha \neq \beta, \beta \neq \gamma, \gamma \neq \alpha\}$$

The number of cases favourable to E is ${}^{26}P_3 = 15,600$.
The number of possible cases in S $= 26^3 = 17,576$. Then

$$P(E) = \frac{15,600}{17,576} = 0.89.$$

Combinations:
The number of unordered sets of distinct elements, i.e. repetition is not allowed.
Examples:

1. Three elements {1,2,3}.
   How many sets (combinations) of two elements from these three?
   Recall for permutations (ordered sequences) there are 6 in all:
   (1,2); (1,3); (2,1); (2, 3); (3,1); (3,2).
   For combinations (1, 2) and (2,1) are equivalent.
   Number of unordered sets = 3:
   {1,2}; {1,3}; {2,3}

2. Four elements {1,2,3,4}.
   How many combinations of two elements from four?
   {1,2}; {1,3}; {1,4} {2,3}; {2, 4}; {3,4};
   Six unordered sets altogether.

3. Four elements {1,2,3,4}.
   How many unordered of three elements from four?
   {1,2,3}; {1,2,4}; {2,3,4}; {3,1,4}
   Four unordered sequences.

Combinations:

Generally:
Number of ways of selecting k distinct elements from n or equivalently number of unordered samples of size k, without replacement from n

$$^nC_k = \frac{^nP_k}{k!} = \frac{n!}{k!(n-k)!}$$

This is the number of combinations of n distinct objects taken k at a time.
Alternative Notation: $\binom{n}{k} = {^nC_k}$.

For example:

1. $^3C_2 = \frac{^3P_2}{2!} = \frac{3*2}{2*1} = 3$
2. $^4C_2 = \frac{^4P_2}{2!} = \frac{4*3}{2*1} = 6$
3. $^4C_3 = \frac{^4P_3}{3!} = \frac{4*3*2}{3*2*1}$

Examples

1. If a box contains 75 good IC chips and 25 defective chips, and 12 chips are selected at random, find the probability that all chips are good.

2. Choose a sample of ten from a class of 100 consisting of 60 females and 40 males.
   What is the probability of getting 10 females?

3. A box with fifteen integrated circuit chips contains five defectives. If a random sample of three chips is drawn, what is the probability that all three are defective?

4. In a party of five persons, compute the probability that at least two have the same birthday (month/day), assuming a 365 day year.

TYPES OF EVENTS

- **UNION OF TWO EVENTS**

  $E_1 \cup E_2$ denotes the outcome of E_1 **or** E_2

  Example: Pull a card from a deck.

  $E_1 =$Spade $\qquad E_2 =$ Ace

  $E_1 \cup E_2 \;=\;$ Ace or Spade

- **INTERSECTION OF TWO EVENTS**

  $E_1 \cap E_2$ denotes the outcome of E_1 **and** E_2

  Example: Pull a card from a deck.

  $E_1 =$Spade $\qquad E_2 =$ Ace

  $E_1 \cap E_2 \;=\;$ Ace and Spade

  $P(E_1 \cap E_2) \;=\;$ P(Ace and Spade)

- **MUTUALLY EXCLUSIVE EVENTS**

  E_1 and E_2 are mutually exclusive if they cannot occur together.

  Example: Pull a card from a deck.

  E_1 =Spade $\qquad E_2$ = Heart

  $E_1 \cap E_2$ is impossible

  $E_1 \cap E_2 = \emptyset$

3 AXIOMS OF PROBABILITY

A probability function P is defined on subsets of the sample space **S** to satisfy the following axioms:

1. Non-Negative Probability:

$$P(E) \geq 0.$$

2. Mutually-Exclusive Events:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

   provided E_1 and E_2 are mutually exclusive.
   i.e. $E_1 \cap E_2$ is empty.

3. The Universal Set:

$$P(S) = 1$$

Example:

Consider the following **if** statement in a program:
if B **then** s_1 **else** s_2
The random experiment consists of "observing" two successive executions of the **if** statement. The sample space consists of the four possible outcomes:

$$S = \{(s_1, s_1), (s_1, s_2), (s_2, s_1), (s_2, s_2)\}$$

Assume that on the basis of strong experimental evidence the following probability assignment is justified:

$P(s_1, s_1) = 0.34,\ P(s_1, s_2) = 0.26,\ P(s_2, s_1) = 0.26,\ P(s_2, s_2) = 0.14,$

Calculate the probability of

1. of at least one execution of the statement s_1
2. that statement s_2 is executed first.

Solution:

1. Let $E =$ At least one execution of the statement s_1

$$E = \{(s_1, s_1),\ (s_1, s_2),\ (s_2, s_1)\}$$

$$P(E_1) = P(s_1, s_1) + P(s_1, s_2) + (s_2, s_1) = 0.86$$

2. Let $E =$ Statement s_2 is executed first.

$$E = \{(s_2, s_1),\ (s_2, s_2)\}$$

$$P(E) = P(s_2, s_1) + P(s_2, s_2) = 0.40$$

Properties of Probability

Theorem 1: Complementary Events

For each $E \subset S$:

$$P(\overline{E}) = 1 - P(E)$$

Proof:

$$S = E \cup \overline{E}$$

Now, E and $\overline{E}$ are mutually exclusive.
i.e.

$$E \cap \overline{E} \text{ is empty.}$$

Hence:

$$P(S) = P(E \cup \overline{E}) = P(E) + P(\overline{E})$$

(Axiom2)

Also:

$$P(S) = 1$$

(Axiom3)

i.e.

$$\begin{aligned} P(S) &= P(E) + P(\overline{E}) \\ \longrightarrow \quad 1 &= P(E) + P(\overline{E}) \end{aligned}$$

So:

$$P(\overline{E}) = 1 - P(E)$$

Properties of Probability

Theorem 2: The Impossible Event/The Empty Set

$$P(\emptyset) = 0 \text{ where } \emptyset \text{ is the empty set}$$

Proof:

$$S = S \cup \emptyset$$

Now: S and $\emptyset$ are mutually exclusive.
i.e.

$$S \cap \emptyset \text{ is empty.}$$

Hence:

$$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$$

(Axiom2)

Also:

$$P(S) = 1$$

(Axiom3)

i.e.

$$1 = 1 + P(\emptyset)$$

i.e.

$$P(\emptyset) = 0.$$

Properties of Probability

Theorem 3:

If E_1 and E_2 are subsets of S such that $E_1 \subset E_2$, then

$$P(E_1) \leq P(E_2)$$

Proof:

$$E_2 = E_1 \cup (\overline{E}_1 \cap E_2)$$

Now, since E_1 and $\overline{E}_1 \cap E_2$ are mutually exclusive,

$$\begin{aligned} P(E_2) &= P(E_1) + P(\overline{E}_1 \cap E_2) \\ &\geq P(E_1) \end{aligned}$$

since $P(\overline{E}_1 \cap E_2) \geq 0$ from Axiom 1.

Properties of Probability

Theorem 4: Range of Probability

For each $E \subset S$

$$0 \leq P(E) \leq 1$$

Proof:
Since,

$$\emptyset \subset E \subset S$$

then from Theorem 3,

$$P(\emptyset) \leq P(E) \leq P(S)$$

$$0 \leq P(E) \leq 1$$

Theorem 5: The Addition Law of Probability

If E_1 and E_2 are subsets of S then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Proof:

$$E_1 \cup E_2 = E_1 \cup (E_2 \cap \overline{E}_1)$$

Now, since E_1 and $E_2 \cap \overline{E}_1$ are mutually exclusive,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2 \cap \overline{E}_1) \qquad (1)$$

(Axiom2)

Now E_2 may be written as two mutually exclusive events as follows:

$$E_2 = (E_2 \cap E_1) \cup (E_2 \cap \overline{E}_1)$$

So

$$P(E_2) = P(E_2 \cap E_1) + P(E_2 \cap \overline{E}_1)$$

(Axiom2)

Thus:

$$P(E_2 \cap \overline{E}_1) = P(E_2) - P(E_2 \cap E_1) \qquad (2)$$

Inserting (2) in (1), we get

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Example:
In a computer installation, 200 programs are written each week, 120 in C^{++} and 80 in Java.
60% of the programs written in C^{++} compile on the first run
80% of the Java programs compile on the first run.

What is the probability that a program chosen at random:

1. is written in C^{++} or compiles on first run?

2. is written in Java or does not compile?

3. either compiles or does not compile?

4 CONDITIONAL PROBABILITY

Example:
In a computer installation, 200 programs are written each week, 120 in C^{++} and 80 in Java. 60% of the programs written in C^{++} compile on the first run and 80% of the Java programs compile on the first run.

	Compiles on first run	Does not compile on first run	
C^{++}	72	48	120
Java	64	16	80
	136	64	200

What is the probability that a program chosen at random:

1. compiles in the first run?

2. is written in C^{++} and compiles on first run?

3. compiles on the first run if we know it has been written in C^{++}?

Conditional Probability

DEFN: The conditional probability of **B** given **A** is

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

Terminology:

- Joint Probability: $P(A \cap B)$
- Marginal Probability: $P(A)$, P(B)
- Conditional Probability: $P(A|B)$ or $P(B|A)$

A rearrangement of the above definition yields the following:

Multiplication Law of Probability:

Two events

$$P(A \cap B) = \begin{matrix} P(A)P(B|A) \\ \\ P(B)P(A|B) \end{matrix}$$

More than two events:

$$P(E_1 \cap E_2 \cap E_3 \cdots \cap E_k) =$$
$$P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \cdots P(E_k|E_1 \cap E_2 \cdots \cap E_{k-1})$$

Examples:

1. Consider four computer firms A, B, C, D bidding for a certain contract. A survey of past bidding success of these firms show the following probabilities of winning:

$$P(A) = 0.35, P(B) = 0.15, P(C) = 0.3, P(D) = 0.2$$

   Before the decision is made to award the contract, firm B withdraws the bid. Find the new probabilities of winning the bid for A, C and D.

2. Pull three cards from a deck without replacement. What is the probability that all are black?

Independent Events

DEFN: Independent Events

A and B are said to be independent if

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

Multiplication Law

Two Independent Events

$$P(A \cap B) = P(A)P(B)$$

More than Two Independent Events

$$P(E_1 \cap E_2 \cap E_3 \cdots E_k) = P(E_1)P(E_2) \cdots P(E_k)$$

Example: Draw 10 cards from a deck with replacement. What is the probability that all are black?

Intel Chips

In October 1994, a flaw was discovered in the Pentium chip installed in many new personal computers. The chip produced an incorrect result when dividing two numbers. Intel, the manufacturer of the Pentium chip, initially announced that such an error would occur once in 9 billion divides, or “once in 27,000 years” for a typical user; consequently it did not immediately offer to replace the chip.

(a) For a division performed using the flawed chip, what is the probability that no error will occur?

(b) Consider two successive divisions performed using the flawed chip. What is the probability that neither result will be in error? (Assume that any one division has no impact on any other division.)

(c) Depending on the procedure, statistical software packages may perform an extremely large number of divisions to produce the required output. For heavy users of the software, one billion divisions over a short time frame is not unusual. Calculate the probability that 1 billion divisions performed using the flawed Pentium chip will result in no errors.

(d) Compute the probability that at least one error occurs in the 1 billion divisions.

Note: Two months after the flaw was discovered, Intel agreed to replace all Pentium chips free of charge.

5 RELIABILITY

Systems Reliability

A system consists of components which determine whether or not it will work. There are various types of configurations of the components in different systems.

- **Series System**
  This is a system in which all the components are in series and they all have to work for the system to work. If one component fails, the system fails.

- **Parallel System**
  This is a system that will fail only if they all fail.

- **Series-Parallel System**
  This is a system where some of the components in series are replicated in parallel.

The Reliability of a System

The reliability of a system is the probability that it is functioning properly. This depends on (i) the reliability of the components and (ii) on the type of system.

Examples:

1. A system consists of 5 components in series each having a reliability of 0.97. What is the reliability of the system?

2. A system consists of 5 components in parallel. Each component has a reliability of 0.97. The system works if at least one of the components works. What is its reliability?

Reliability with series systems
The problem with series systems is that reliability quickly decreases as the number of components increases.

Reliability with Parallel Systems
The problem with parallel systems is that the 'law of diminishing returns' operates. The rate of increase in reliability with each additional component decreases as the number of components increases.

Most systems are combinations of series and parallel systems

Example: A series parallel system:

Consider a system with 5 components.
The reliability of

1. component 1 is 0.95,
2. component 2 is 0.95,
3. component 3 is 0.7,
4. component 4 is 0.75
5. component 5 is 0.90.

Because of the low reliability of the third and fourth components, they are replicated; the third component is replicated 3 times and the 4th component is replicated twice. Calculate the overall reliability of the system.

6 BAYES THEOREM

Law of Total Probability
If a sample space can be partitioned into k mutually exclusive and exhaustive events:

$$A_1, A_2, A_3, \cdots A_k$$

i.e.

$$S = A_1 \cup A_2 \cup A_3 \cdots \cup A_k$$

Then for any event E:

$$P(E) = P(A_1)P(E|A_1) + P(A_2)(PE|A_2) \cdots P(A_k)P(E|A_K)$$

Proof:

$$\begin{aligned} E &= E \cap S \\ &= E \cap (A_1 \cup A_2 \cup \cdots \cup A_k) \\ &= (E \cap A_1) \cup (E \cap A_2) \cup \cdots (E \cap A_k) \end{aligned}$$

Since these are mutually exclusive

$$\begin{aligned} P(E) &= P(E \cap A_1) + P(E \cap A_2) + \cdots P(E \cap A_k) \\ &= P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + \cdots P(A_k)P(E|A_k) \end{aligned}$$

Examples:

1. In a computer installation, 60% of programs are written in C^{++} and 40% in Java. 60% of the programs written in C^{++} compile on the first run and 80% of the Java programs compile on the first run.

   (a) What is the overall proportion of programs that compile on first run?

   (b) If a randomly selected program compiles on the first run what is the probability that it was written in C^{++}?

2. In a certain company

   50% of documents are written in WORD; 30% in LATEX; 20% in HTML.
   From past experience it is know that :
   40% of the WORD documents exceed 10 pages
   20% of the LATEX documents exceed 10 pages
   20% of the HTML exceed 10 pages

   (a) What is the overall proportion of documents containing more than 10 pages?

   (b) A document is chosen at random and found to to have more than 10 pages. What is the probability that it has been written in LATEX?

Examples:

3. Enquiries to an on-line computer system arrive on 5 communication lines. The percentage of messages received through each line are:

Line	1	2	3	4	5
% received	20	30	10	15	25

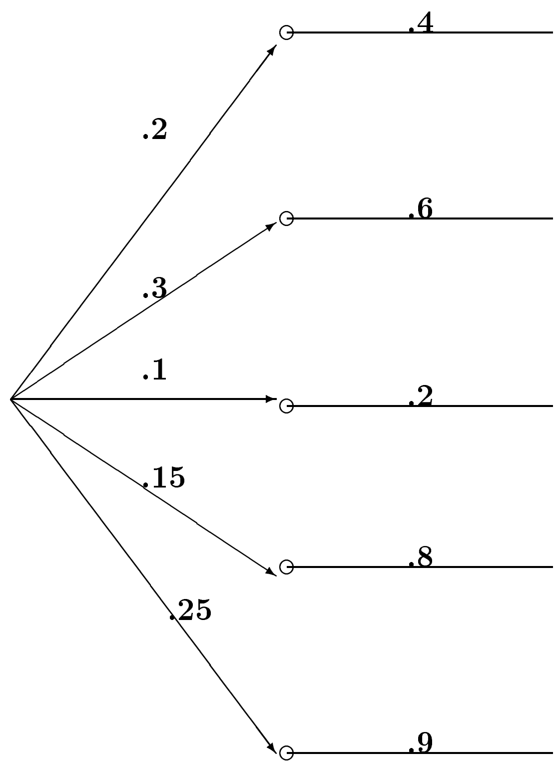
From past experience, it is known that the percentage of messages exceeding 100 characters on the different lines are:

Line	1	2	3	4	5
% exceeding 100 characters	40	60	20	80	90

(a) Calculate the overall proportion of messages exceeding 100 characters.

(b) If a message chosen at random is found to exceed 100 characters, what is the probability that it came through line 5?

Tree Diagram



Examples:

4. A binary communication channel carries data as one of two sets of signals denoted by 0 and 1. Owing to noise, a transmitted 0 is sometimes received as a 1, and a transmitted 1 is sometimes received as a 0. For a given channel, it can be assumed that a transmitted 0 is correctly received with probability 0.95 and a transmitted 1 is correctly received with probability 0.75. Also, 60% of all messages are transmitted as a 0. If a signal is sent, determine the probability that:

   (a) a 1 was received;

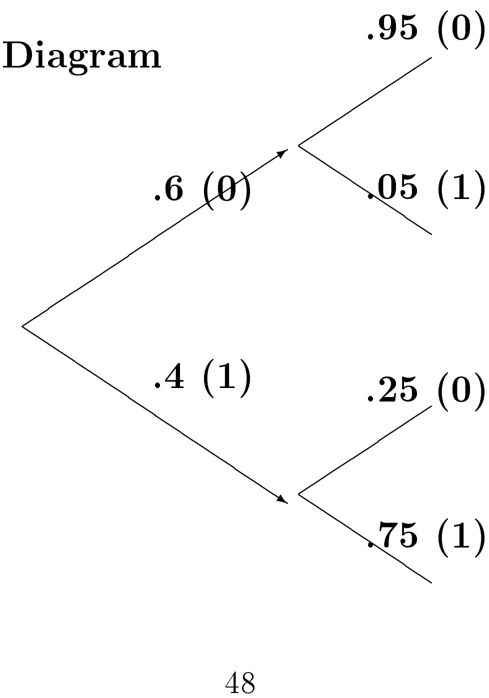
   (b) a 0 was received;

   (c) a 1 was transmitted given than a 1 was received;

   (d) a 0 was transmitted given that a 0 was received;

   (e) an error occurred.

Tree Diagram



Bayes' Theorem

If a sample space can be partitioned into k mutually exclusive and exhaustive events:

$$A_1, A_2, A_3, \cdots A_k$$

$$S = A_1 \cup A_2 \cup A_3 \cdots \cup A_k$$

Then for any event E,:

$$P(E) = P(A_1)P(E|A_1) + P(A_2)P(E|A_2) \cdots P(A_k)P(E|A_k)$$

$$P(A_i|E) = \frac{P(A_i)P(E|A_i)}{P(E)}$$

Proof: For any $i,\ 1 \leq i \leq k$

$$\begin{aligned} E \cap A_i &= A_i \cap E \\ P(E)P(A_i|E) &= P(A_i)P(E|A_i) \\ P(A_i|E) &= \frac{P(A_i)P(E|A_i)}{P(E)} \end{aligned}$$

$P(A_i|E)$ is called the **POSTERIOR PROBABILITY**

Example:

A software company surveyed office managers to determine the probability that they would buy a new graphics package. Eighty percent of the office managers claimed that they would buy the package. Of those managers who would buy the graphics package, 40% were also interested in upgrading their computer hardware. Of those managers who were not interested in purchasing the graphics package, only 10% were interested in upgrading their computer hardware.

(a) What is the probability that an office manager who is interested in upgrading her computer hardware is also interested in purchasing the graphics package?

7 RANDOM VARIABLES

Defn: A random variable is a rule which assigns a numerical value to each possible outcome of an experiment

Example: Toss a coin: S = H, T
Call a head 1 and a tail 0
S = { 1, 0 }

Random variables are **DISCRETE** or **CONTINUOUS**

- **Discrete Random Variable**:
  A random variable is discrete if its values can assume isolated points on the number line.
  Examples:

  number of telephone calls in an hour

  number of jobs arriving for service

- **Continuous Random Variable:**
  A random variable is continuous if its values can assume all points in a particular interval.

  Examples:

  Heights, Weights;

  Lifetime of a component;

  cpu time;

  response time

Probability Distributions:

Defn: The probability distribution consists of all possible values of a variable and its associated probabilities.

Examples:

1. Toss a coin:

X = x	0	1
P(X = x)	0.5	0.5

2. Toss a coin twice

$(X = x, Y = y)$	(0,0)	(0, 1)	(1,0)	(1,1)
P$(X = x \cap Y = y)$	0.25	0.25	0.25	0.25

3. Roll a die:

X = x	1	2	3	4	5	6
P(X = x)	1/6	1/6	1/6	1/6	1/6	1/6

4. The number of hardware failures of a computer system in a week has the following probabilities:

No.of Failures	0	1	2	3	4	5	6
Probability	.18	.28	.25	.18	.06	.04	.01

Probability Distributions

Probability Density Function (PDF) satisfies

- $P(X = x) \geq 0$
- $\sum_x P(X = x) = 1$

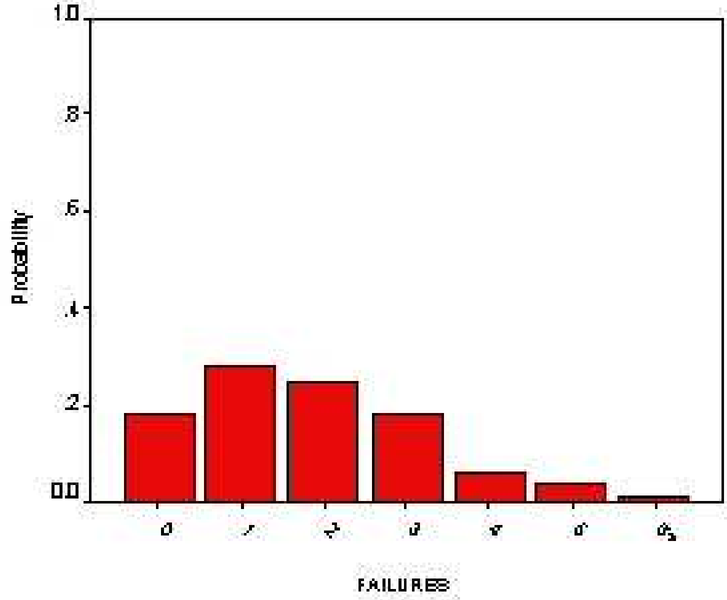
Cumulative Distribution Function (CDF):

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

Example: Hardware Failures

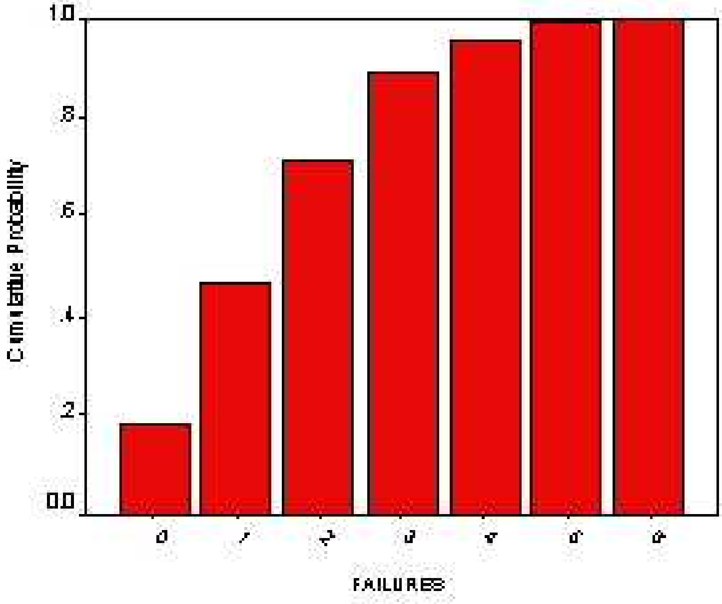
No.of Failures	0	1	2	3	4	5	6
PDF P(X=x)	.18	.28	.25	.18	.06	.04	.01
CDF P($X \leq x$)	.18	.46	.71	.89	.95	.99	1.00

Figure 3: **Probability Density Function**



No.of Failures	0	1	2	3	4	5	6
PDF P(X=x)	.18	.28	.25	.18	.06	.04	.01

Figure 4: **Cumulative Distribution Function**



No.of Failures	0	1	2	3	4	5	6
CDF $P(X \leq x)$	.18	.46	.71	.89	.95	.99	1.00

8 GEOMETRIC DISTRIBUTION

EXAMPLES:

1. Toss a coin repeatedly.
   Let X = number of tosses to first head

2. It is known that 20% of products on a production line are defective. Products are inspected until first defective is encountered.
   Let X = number of inspections to obtain first defective

3. Observing single births until a girl.
   Let X = number of observations to first girl.

4. Terminals on an on-line computer system are attached to a communication line to the central computer system. The probability that any terminal is ready to transmit is 0.95.
   Let X = number of terminals polled until the first ready terminal is located.

GEOMETRIC DISTRIBUTION
Conditions:

1. An experiment consists of repeating trials until first success.
2. Each trial has two possible outcomes;
   (a) A success with probability p
   (b) A failure with probability $q = 1 - p$.
3. Repeated trials are independent.

X = number of trials to first success

X is a **GEOMETRIC RANDOM VARIABLE**.

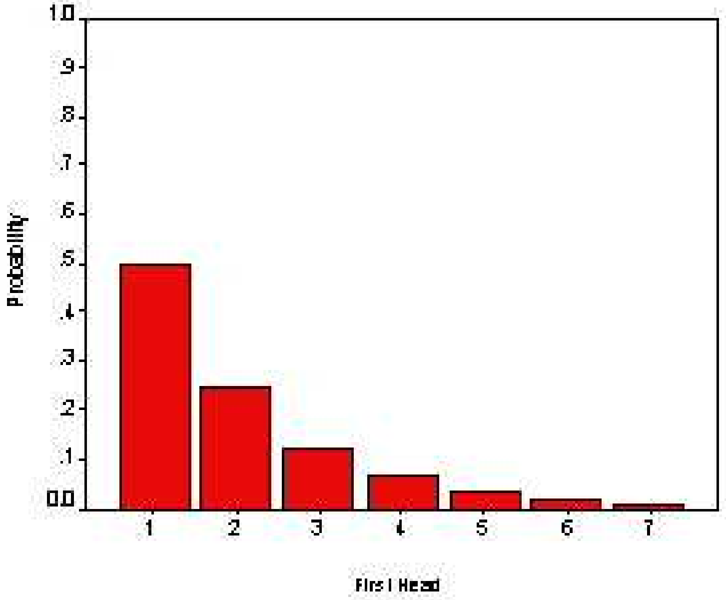
PDF:

$$P(X = x) = q^{x-1}p; \quad x = 1, 2, 3, \cdots$$

CDF:

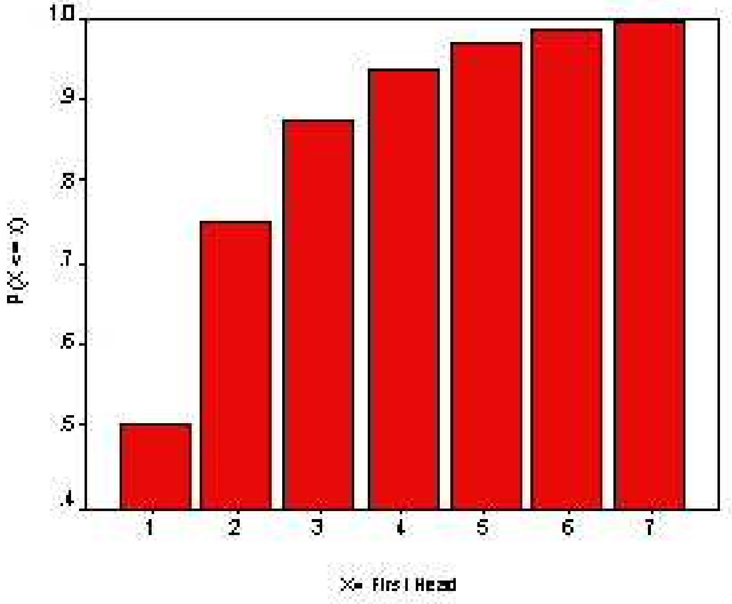
$$\begin{aligned} P(X \le x) &= P(X = 1) + P(X = 2) \cdots P(X = x) \\ &= p + qp + q^2 p \cdots + q^{x-1} p \\ &= p[1 - q^x]/(1 - q) \\ &= 1 - q^x \end{aligned}$$

Figure 5: Probability to First Head



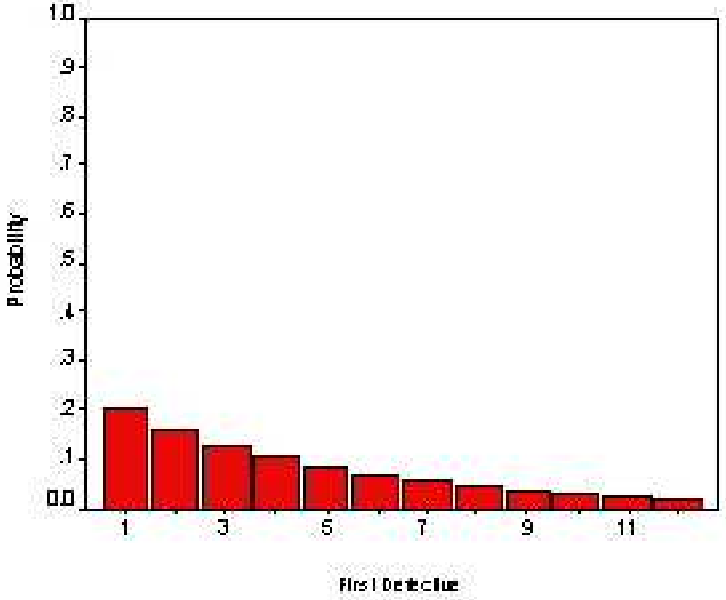
First Head	1	2	3	4	5	6	7	8	9	10
Probability	.5	.25	.125	.0625	.0313	.0156	.0078	.0039	.0020	.0010

Figure 6: Cumulative Distribution Function



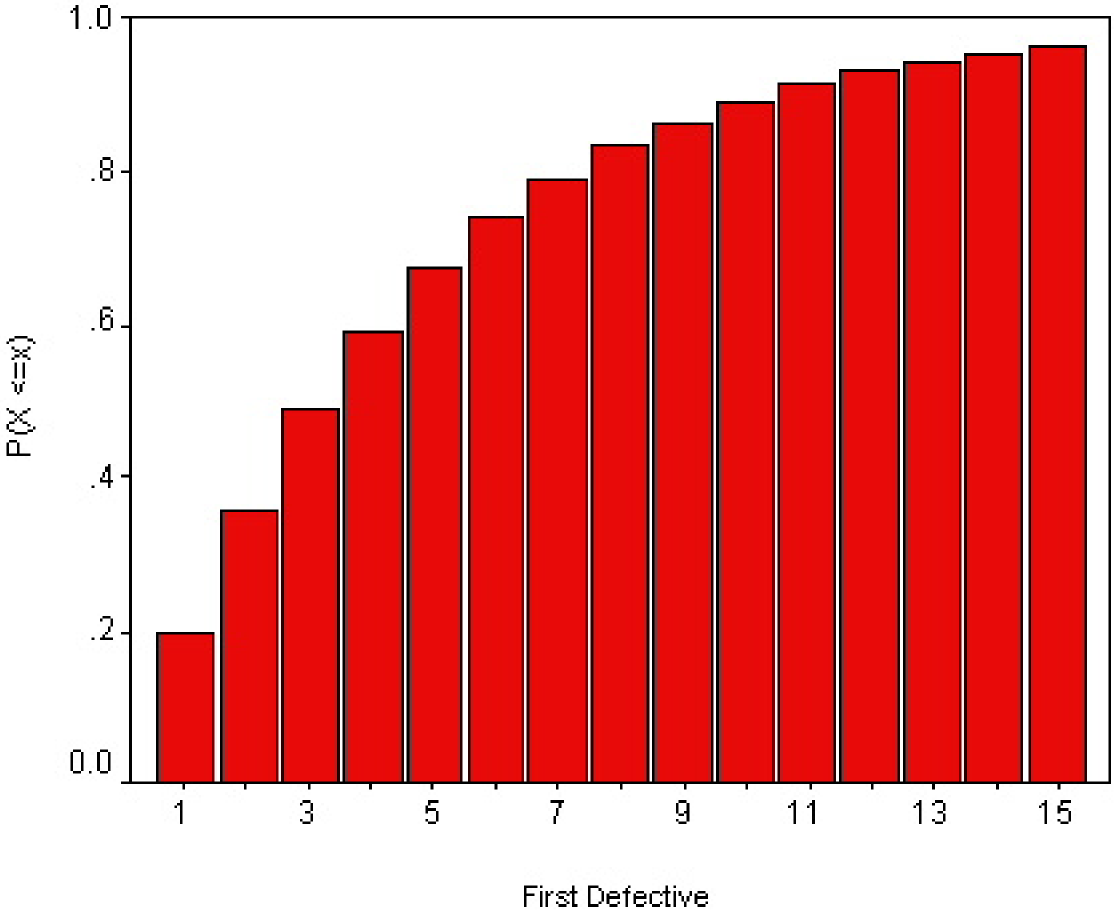
First Defective	1	2	3	4	5	6	7	8	9
Probability	.5	.75	.875	.09375	.9688	.9844	.9922	.9961	.9980.9999

Figure 7: Probability to First Defective



First Defective	1	2	3	4	5	6	7	8	9	1
Probability	.2	.16	.128	.1024	.0819	.0655	.0524	.0419	.0338	.026

Figure 8: Probability to First Defective



First Defective	1	2	3	4	5	6	7	8	9	10	$\cdots$
$P(X \leq x)$	.2	.36	.488	.5904	.6723	.7379	.7903	.8322	.8685	.8926	$\cdots$

Example:
Products produced by a machine has a 3% defective rate.

- What is the probability that the first defective occurs in the fifth item inspected?

$$\begin{aligned} P(X=5) &= P(\text{1st 4 non-defective })P(\text{ 5th defective}) \\ &= \left(0.97^4\right)(0.03) \end{aligned}$$

- What is the probability that the first defective occurs in the first five inspections?

$$\begin{aligned} P(X \leq 5) &= 1 - P(\text{First 5 non-defective}) \\ &= 1 - 0.97^5 \end{aligned}$$

The Markov Property:

If the probability of events happening in the future is independent of what went before, then the random variable is said to have the **Markov property**.

i.e

$$P(X = n + x | X > n) = P(X = x)$$

MARKOV PROPERTY

$\Longrightarrow$ MEMORYLESS PROPERTY

Example: Markov Property

Products are inspected until first defective is found. X is a geometric random variable with parameter p. The first 10 trials have been found to be free of defectives. What is the probability that the first defective will occur in the 15th trial?

Let E_1 be the event that first ten trials are free of defectives.

Let E_2 be the event that that first defective will occur on the 15th trial.

$$\begin{aligned} P(X = 15|X > 10) &= P(E_2|E_1) \\ &= \frac{P(E_1 \cap E_2)}{P(E_1)} \\ &= \frac{P(X = 15 \cap X > 10)}{P(X > 10)} \\ &= \frac{P(X = 15)}{P(X > 10)} \\ &= \frac{q^{14}p}{q^{10}} \\ &= q^4 p \\ &= P(X = 5) \end{aligned}$$

MARKOV PROPERTY

Generally, the Markov property states:

$$P(X = x + n|X > n) = P(X = x)$$

Proof:
Let

$$\begin{aligned} E_1 &= \{X > n\} \\ E_2 &= \{X = x + n\} \end{aligned}$$

Then we may write

$$P(X = x + n|X > n) = P(E_2|E_1)$$

But

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

Now

$$P(E_1 \cap E_2) = P(X = x + n) = q^{x+n-1}p$$

And

$$P(E_1) = P(X > n) = q^n$$

Thus

$$\begin{aligned} P(E_2|E_1) &= \frac{q^{x+n-1}p}{q^n} \\ &= q^{x-1}p \end{aligned}$$

But

$$P(X = x) = q^{x-1}p$$

Hence

$$P(X = x + n|(X > n) = P(X = x)$$

9 BINOMIAL DISTRIBUTION

Examples:

1. Toss a coin 6 times
   Let X = number of heads

2. A production line is known to produce 10% defective.
   Choose 10 products for inspection
   Let X = number of defectives

3. A fair die is rolled 5 times.
   Let X = number of 'twos'

4. A player scores on 75% of frees.
   Let X = number of scores in four frees.

5. A printer in a computer lab. works 60% of the time.
   Let X = number of times the printer is working in 10 visits.

6. Five terminals on an on-line computer system are attached to a communication line to the central computer system. The probability that any terminal is busy is .1.
   Let X = number of busy terminals.

BINOMIAL CONDITIONS

1. An experiment consists of n repeated trials.
2. Each trial has two possible outcomes: success or failure.
3. The probability of a success p is constant from trial to trial.
4. Repeated trials are independent.

Let X = number of successes in n trials

X is a BINOMIAL random variable.

General Binomial Distribution

n = no of trials

p = probability of success

$q = 1 - p$ = probability of failure

X = no of successes in n trials

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Example: Busy terminals n = 5, p = .1, q= .9

X	0	1	2	3	4	5
PDF	0.591	0.328	0.073	0.008	0.001	0.000
CDF	0.591	0.919	0.992	0.999	1.000	1.000

Figure 9: Probability Density Function

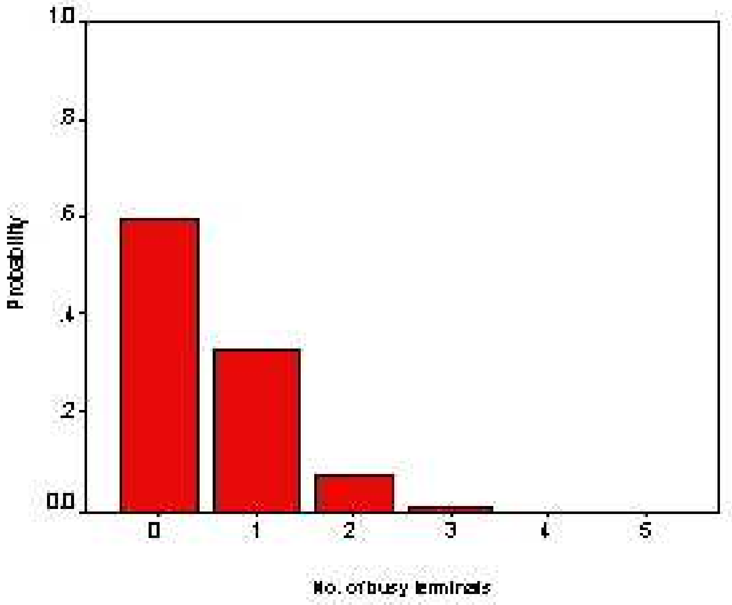
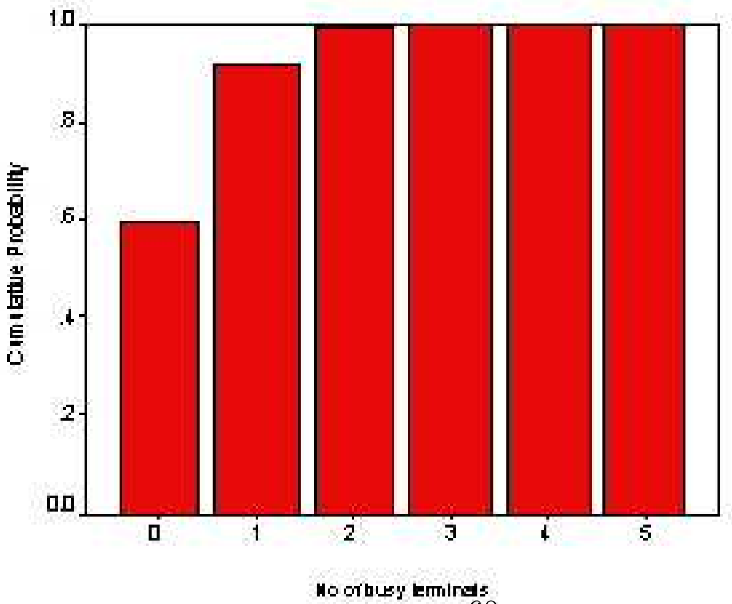


Figure 10: Cumulative Distribution Function



Example: Defectives n = 10, p = .1, q= .9

X	0	1	2	3	4	5	6
PDF	0.3486	.3874	.1937	.057	.0112	.0015	.0001
CDF	0.3486	.7361	.9298	.9872	.9984	.9999	1.000

Figure 11: Probability Density Function

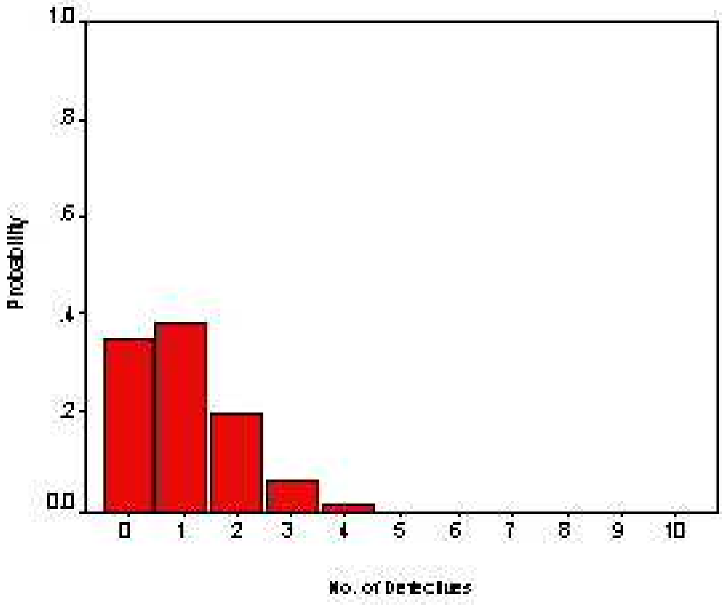
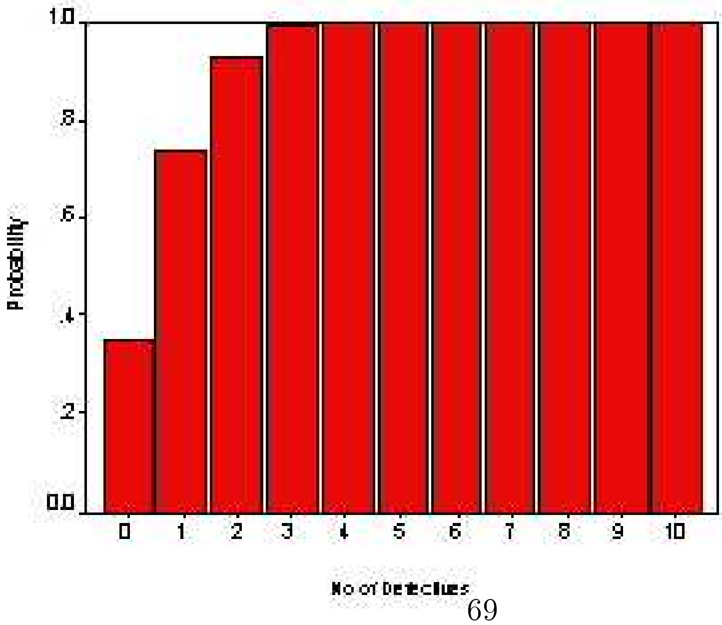


Figure 12: Cumulative Distribution Function



EXAMPLE :

Based on past experience, a printer in a laboratory is operating 60% of the time. Throughout a particular day, 8 visits are made and the number of times, X, that the printer is operating is observed.

1. Write the PDF and CDF of X

2. What is the probability that the printer is operating:

   (a) exactly 4 times?

   (b) at least 4 times?

   (c) at most 4 times?

   (d) more than 4 times?

   (e) fewer than 4 times?

10 SAMPLING WITHOUT REPLACEMENT

Example:
A box contains 20 items of which 10% are defective.
A sample of 10 is drawn at random.
What is the probability that the sample will contain 2 defectives?
Solution:

1. Sampling with Replacement $\Longrightarrow$ Binomial
   Binomial with $n = 10$ and $p = 0.1$; Bino(10, 0.1)

$$P(\ 2\ \text{ defectives}) = \begin{pmatrix} 10 \\ 2 \end{pmatrix} .1^2 .9^8 = 0.1937$$

2. Sampling without Replacement

2 Defectives	18 Non-defectives

$$P(2\ defectives) = \frac{\begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{pmatrix} 18 \\ 8 \end{pmatrix}}{\begin{pmatrix} 20 \\ 10 \end{pmatrix}} = 0.2368$$

Hypergeometric Distribution

A finite population of size N consists of

M elements called successes
L elements called failures

A sample of n elements are selected at random **without replacement**.
X = number of successes

$$P(X = x) = \frac{\begin{pmatrix} M \\ x \end{pmatrix} \begin{pmatrix} L \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

X is said to have a **hypergeometric distribution**

Example:
Draw 6 cards from a deck without replacement.
What is the probability of getting two hearts?
Solution:
Here

$$\begin{aligned} M &= 13 \quad \text{number of hearts} \\ L &= 39 \quad \text{number of non-hearts} \\ N &= 52 \quad \text{total} \end{aligned}$$

$$P(2\ hearts) = \frac{\begin{pmatrix} 13 \\ 2 \end{pmatrix} \begin{pmatrix} 39 \\ 4 \end{pmatrix}}{\begin{pmatrix} 52 \\ 6 \end{pmatrix}} = .31513$$

Example: Lotto
42 balls are numbered 1 - 42.
You select six numbers between 1 and 42. (The ones you write on your lotto card)
Number of possible ways to draw six mumbers in the range $[\ 1,\, 42\ ] = \binom{42}{6}$

What is the probability that they contain

(i) match 6?

(ii) match 5?

(ii) match 4?

(iii) match 3?

Solution:
Total = 42; Favourable = 6; Non-Favourable = 36.

Sample size n = 6.

$$P(\text{match } \ 4) = \frac{\binom{6}{4}\binom{36}{2}}{\binom{42}{6}} = .0018$$

ODDS OF ABOUT 1 in 500

Binomial or Hypergeometric?

Boxes contain 20 items of which 10% are defective.

Find the probability that no more than 2 defectives will be obtained in a sample of size 10.

Let X = no of defectives.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \quad = .3487 \\ P(X = 1) &= \binom{10}{1} .1^1 .9^9 = .3874 \\ P(X = 2) &= \binom{10}{2} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$\begin{aligned} P(X = 0) &= \frac{\binom{18}{10}}{\binom{20}{10}} = .2368 \\ P(X = 1) &= \frac{\binom{2}{1}\binom{18}{9}}{\binom{20}{10}} = .5263 \\ P(X = 2) &= \frac{\binom{2}{2}\binom{18}{8}}{\binom{20}{10}} = .2368 \end{aligned}$$

Binomial or Hypergeometric?
Boxes contain 200 items of which 10% are defective.
Find the probability that no more than 2 defectives will be obtained in a sample of size 10.
Let X = no of defectives.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \quad = .3487 \\ P(X = 1) &= \binom{10}{1} .1^1 .9^9 = .3874 \\ P(X = 2) &= \binom{10}{2} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$\begin{aligned} P(X = 0) &= \frac{\binom{180}{10}}{\binom{200}{10}} = .3398 \\ P(X = 1) &= \frac{\binom{20}{1}\binom{180}{9}}{\binom{200}{10}} = .3974 \\ P(X = 2) &= \frac{\binom{20}{2}\binom{180}{8}}{\binom{200}{10}} = .1975 \end{aligned}$$

Binomial or Hypergeometric?
Boxes contain 2000 items of which 10% are defective.
Find the probability that no more than 2 defectives will be obtained in a sample of size 10.
Let X = no of defectives.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \quad = .3487 \\ P(X = 1) &= \binom{10}{1} .1^1 .9^9 = .3874 \\ P(X = 2) &= \binom{10}{2} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$\begin{aligned} P(X = 0) &= \frac{\binom{1800}{10}}{\binom{2000}{10}} = .3476 \\ P(X = 1) &= \frac{\binom{200}{1}\binom{1800}{9}}{\binom{2000}{10}} = .3881 \\ P(X = 2) &= \frac{\binom{200}{2}\binom{1800}{8}}{\binom{2000}{10}} = .1939 \end{aligned}$$

Binomial or Hypergeometric?

What is the probability of getting no more than 2 defectives in a random sample drawn without replacement from a batch which has 10% defectives?

Batch Size	20	200	2000	Bin. Approx
P(X=0)	.2368	.3398	.3476	.3487
P(X=1)	.5263	.3974	.3881	.3874
P(X=2)	.2368	.1975	.1939	.1937
P(X≤ 2)	.999	.9347	.9296	.9298

Theorem: As $N \to \infty$, the hypergeometric distribution converges to the binomial.

Proof:

$$\begin{aligned}
\text{Population Size} &= N \\
\text{Proportion of successes} &= p \\
\text{Number of successes in N} &= Np \\
\text{Number of failures} &= N(1-p)
\end{aligned}$$

Let X = number of successes in s sample of size n drawn without replacement from N

Np Successes	N(1-p) Failures

We will show that for the hypergeometric

$$\begin{aligned}
P(X = x) &= \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{x}} \\
&\to \binom{n}{x} p^x q^{n-x} \quad \text{as} \quad N \to \infty
\end{aligned}$$

Proof:

$$
\begin{aligned}
P(X=x) &= \frac{\begin{pmatrix} Np \\ x \end{pmatrix}\begin{pmatrix} Nq \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
&= \frac{\left(\dfrac{(Np)!}{x!(Np-x)!}\right)\left(\dfrac{(Nq)!}{(n-x)!\,(Nq-(n-x))!}\right)}{\dfrac{N!}{n!(N-n)!}} \\
&= \frac{n!}{x!(n-x)!}\left[\frac{(Np)!}{(Np-x)!}\frac{(Nq)!}{(Nq-(n-x))!}\frac{(N-n)!}{N!}\right] \\
&\to \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x} \quad \text{as} \quad N \to \infty
\end{aligned}
$$

$$\begin{aligned}\frac{(Np)!}{(Np-x)!} &= Np(Np-1)(Np-2)\cdots(Np-x+1) \\ &= N^x p(p-\frac{1}{N})(p-\frac{2}{N})\cdots(p-\frac{x-1}{N})\end{aligned}$$

$$\begin{aligned}\frac{Nq}{(Nq-(n-x))!} &= (Nq)(Nq-1)\cdots(Nq-(n-x)+1) \\ &= N^{n-x}q(q-\frac{1}{N})\cdots(q-\frac{n-x+1}{N})\end{aligned}$$

$$\begin{aligned}\frac{(N-n)!}{N!} &= \frac{1}{N(N-1)(N-2)\cdots(N-n+1)} \\ &= 1/\left(N^n(1-\frac{1}{N})(1-\frac{2}{N})\cdots(1-\frac{n-1}{N})\right)\end{aligned}$$

11 EXPECTATIONS

The Mean of a Sample

Example 1:
Salaries of 6 recent computer science graduates (£000s).

$$20.3, 14.9, 18.9, 21.7, 16.3, 17.7$$

The average of the sample is

$$\overline{x} = \frac{20.3 + 14.9 + 18.9 + 21.7 + 16.3 + 17.7}{6} = 18.3$$

Generally, if $x_1, x_2, \cdots x_n$ in a sample of size n,

Then the sample average is

$$\overline{x} = \frac{x_1 + x_2 + \cdots x_n}{n}$$

It may be written as

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example 2: Averages

One hundred applicants for a certain degree program had the follwing age distribution.

Age	No. of applicants
18	9
19	40
20	18
21	18
22	8
23	4
24	3

The mean is obtained as follows:

$$\overline{x} = \frac{(18*9)+(19*40)+(20*18)+(21*18)+(22*8)+(23*4)+(24*3)}{100} =$$

$$\overline{x} = (18*.09)+(19*.40)+(20*.18)+(21*.18)+(22*.08)(23*.04)(24*.03) = 20$$

$$\sum xp(x)$$

Mean of a Random Variable

Definition: The mean of a discrete random variable is defined as the weighted average of all possible values. The weights are the probabilities of respective values of the random variable

$$E(X) = \sum_x xp(x)$$

The mean or expected value of a random variable X is often denoted by μ_x

Examples:

1. The number of hardware failures of a computer system in a week of operation has the following pdf:

No. of Failures (X)	0	1	2	3	4	5	6
Probability (P(X=x))	.18	.28	.25	.18	.06	.04	.01

   Calculate the expected number of failures in a week.

2. A quarter of the source programs submitted by a certain programmer compile successfully. Each day the programmer writes five programs. The compiling probabilities are:

No. that compiles	0	1	2	3	4	5
Probability	.237	.396	264	.088	.014	.001

   Calculate the average number of programs that compile in a day.

Some Derivations:

- **Mean of binomial distribution**

$$PDF : p(x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \ldots, n$$

$$\begin{aligned}
MEAN &: \\
E(X) &= \sum_x x p(x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x q^{n-x} \\
&= np \sum_{x-1=0}^{n-1} \binom{n-1}{x-1} p^{x-1} q^{(n-1)-(x-1)} \\
&= np \, [p+q]^{n-1} \\
&= np
\end{aligned}$$

- **Mean of geometric distribution**:
  Example: Terminals on an on-line computer system are attached to a communication line to the central computer system. The probability that any terminal is ready to transmit is 0.95.

$$
\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x q^{x-1} p \\
&= p \sum_{x=1}^{\infty} x q^{x-1} \\
&= p \sum_{x=1}^{\infty} \frac{dq^x}{dq} \\
&= p \frac{d \sum_{x=1}^{\infty} q^x}{dq} \\
&= p \frac{d(q/(1-q)}{dq} \\
&= p \frac{[(1-q)+q]}{(1-q)^2} \\
&= \frac{p}{p^2} = \frac{1}{p}
\end{aligned}
$$

12 VARIANCES

Variance of Sample:

Spread of individual values from the mean.

Example 1:
Salaries of 6 recent computer science graduates (£000)

$$20.3, 14.9, 18.9, 21.7, 16.3, 17.7$$

Recall $\overline{x} = 18.3$.

Calculation of s^2:

$$s^2 = \frac{\left[(20.3 - 18.3)^2 + (18.9 - 18.3)^2 + \ldots\right]}{5}$$

Generally, if $x_1, x_2, \cdots x_n$ is a sample of size n,

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$s^2 = \frac{\left[(20.3 - 18.3)^2 + (18.9 - 18.3)^2 + \ldots\right]}{5} = 6.368$$

Standard Deviation:

$$s = \sqrt{6.368} = 2.523$$

Example 2: Variance

One hundred applicants for a certain degree program had the follwing age distribution.

Age	No. of applicants
18	9
19	40
20	18
21	18
22	8
23	4
24	3

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 f_i}{\sum_i f_i}$$

The variance is obtained as follows:

$$\overline{x} = (18-20)^2 * .09 + (19-20)^2 * .40 + (20-20)^2 * .18 + (21-20)^2 * .18$$
$$+(22-20)^2 * .08 + (23-20)^2 * .04 + (24-20)^2 * .03$$

Check

$$\sum(x-20)^2 p(x)$$

$$s^2 = 2.1$$

The variance of a discrete random variable

Definition: The variance is defined as the weighted average of the squared differences between each possible outcome and its mean ... the weights being the probability of the respective outcomes

$$V(X) = \sum_x (x - \mu_x)^2 p(x)$$

Or equivalently

$$V(X) = E\left(X - (E(X)\right)^2$$

The variance is often denoted by σ_x^2.

$$\sigma_x^2 \equiv E(X - \mu_x)^2$$

The standard deviation is denoted by σ_x.

$$\sigma_x \; = \sqrt{(E(X - \mu_x)^2)}$$

Examples:

1. A quarter of the source programs submitted by a certain programmer compile successfully. Each day the programmer writes five programs. The compiling probabilities are:

No. that compiles	0	1	2	3	4	5
Probability	.237	.396	.264	.088	.014	.001

Calculate:

(a) The expected number of programs that will compile per day

$$E(X) = (0*.237)+(1*.396)+(2*.264)+(3*.088)+(4*.014)+(5*.001)$$
$$= 1.25$$

(b) The variance.

X	0	1	2	3	4	5
P(X=x)	.237	.396	.264	.088	.014	.001
$(x-1.25)^2$	1.5625	.0625	.5625	3.0625	7.5625	14.0625
$(x-1.25)^2 p(x)$	.3703	.0247	.1485	.2695	.1059	.0140

$$V(X) = \sum_x (x-\mu_x)^2 p(x) = .933$$

Some Derivations:
Variance of a binomial distribution

$$V(X) = E(X - \mu_x)^2 = \sum_x (x - \mu_x)^2 p(x)$$

Recall: the pdf of the binomial distribution:

$$\begin{aligned} p(x) &= \binom{n}{x} p^x q^{n-x}, \ x = 0, 1, \ldots, n \\ \mu_x &= np \end{aligned}$$

Show $V(X) = npq$
Proof:

$$\begin{aligned} V(X) &= \sum_{x=0}^{n} (x - np)^2 \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^{n} \left[x^2 - 2npx + (np)^2\right] \binom{n}{x} p^x q^{n-x} \end{aligned}$$

$$\begin{aligned} V(X) &= \\ & \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} - 2np \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} + n^2 p^2 \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} \\ &= (n^2p^2 - np^2 + np) - 2n^2p^2 + n^2p^2 \\ &= np - np^2 \\ &= np(1 - p) \end{aligned}$$

$$
\begin{aligned}
\sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} + \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} + np \\
&= \sum_{x=2}^{n} n(n-1) \binom{n-2}{x-2} p^x q^{n-x} + np \\
&= n(n-1)p^2 \sum_{x-2=0}^{n-2} \binom{n-2}{x-2} p^{x-2} q^{(n-2)-(x-2)} + np \\
&= n(n-1)p^2 \left[p+q\right]^{n-2} + np \\
&= n^2p^2 - np^2 + np
\end{aligned}
$$

13 PROPERTIES OF EXPECTATIONS

RECALL

$$E(X) = \sum_x xp(x) \text{ when } X \text{ is discrete}$$

Example
The average salary of employees in a computer firm is £27,500. After negotiations with the trade union, it was agreed that employees would get a rise of £100 in addition to 10 percent increase on their basic salaries. What is the new average salary?

Let X = old salary; Y = new salary.

$$Y = 100 + 1.1X$$

If E(X) = £27,500, what is E(Y)?

Properties of Expected Values

1. E(X+c) = E(X) + c

$$
\begin{aligned}
E(X+c) &= \sum_x (x+c)p(x) \\
&= \sum_x xp(x) + c\sum_x p(x) \\
&= \; = E(X) + c.
\end{aligned}
$$

2. E(cX)= cE(X)

$$
\begin{aligned}
E(cX) &= \sum_x cxp(x) \\
&= c\sum_x xp(x) \\
&= cE(X).
\end{aligned}
$$

3. V(X+c) = V(X)

$$\begin{aligned} V(X+c) &= E[X+c-E(X+c)]^2 \\ &= E[X-E(X)]^2 \\ &= V(X) \end{aligned}$$

4. $V(cX) = c^2 V(X)$

$$\begin{aligned} V(cX) &= E[cX-E(cX)]^2 \\ &= E[cX-cE(X)]^2 \\ &= E[c(X-E(X))]^2 \\ &= c^2 E[X-E(X)]^2 \\ &= c^2 V(X) \end{aligned}$$

14 POISSON DISTRIBUTION

Examples

1. Number of telephone calls in a 15-minute interval.

2. Number of people arriving at a checkout in an hour.

3. Number of jobs arriving for service in a day.

Generally
X = number of events, distributed independently in time, occurring in a fixed time interval.

X is a Poisson variable with pdf:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots, \infty$$

where λ is the average.

Example:

Consider a computer system with Poisson job-arrival stream at an average of 2 per minute. Determine the probability that in any one-minute interval there will be (i) 0 jobs; (ii) exactly 2 jobs; (iii) more than 3 arrivals.

Figure 13: Poisson PDF $\lambda = 2$

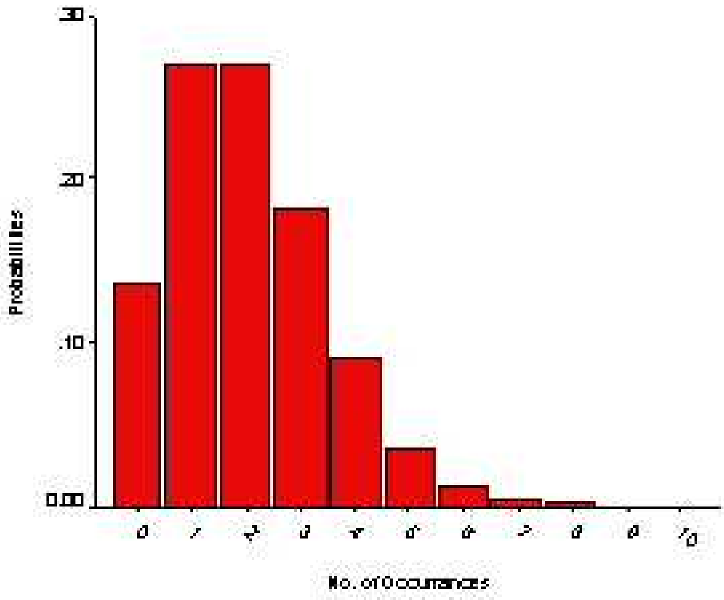
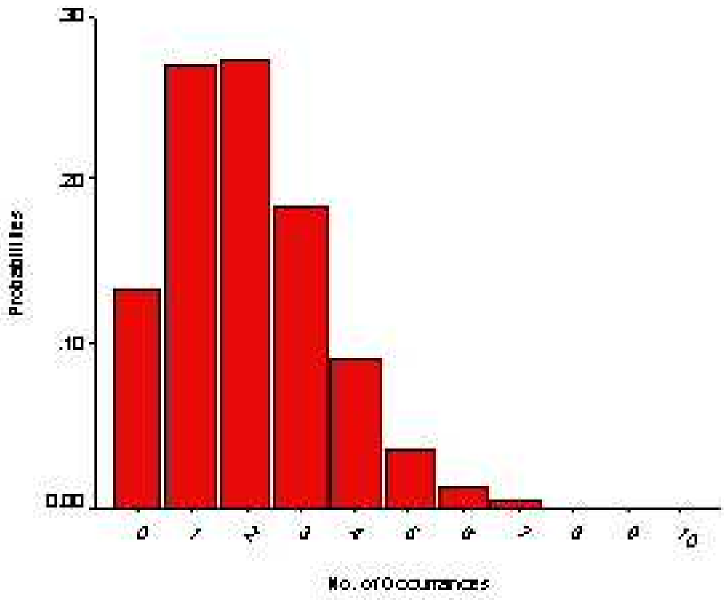


Figure 14: Binomial PDF $n = 100$ and $p = .02$



Derivations of Some Properties of Poisson

1. Clearly

$$e^{-\lambda}\frac{\lambda^x}{x!} > 0 \quad \text{since} \quad \lambda > 0$$

Also

$$\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = 1$$

since

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots$$

i.e.

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$$

2. $E(X) = \lambda$

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x e^{-\lambda}\frac{\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \\
&= e^{-\lambda}\lambda \left[\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right] \\
&= e^{-\lambda}\lambda \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots \right] \\
&= e^{-\lambda}\lambda e^{\lambda} = \lambda
\end{aligned}$$

15 APPLICATIONS OF THE POISSON

The Poisson distribution arises in two ways:

1. **As an approximation to the binomial when p is small and n is large**

   Example:

   When examining the number defectives in a large batch. p, the defective rate, is usually small.

2. **Events distributed independently of one another in time:**
   X = the number of events occurring in a fixed time interval has a Poisson distribution.

   Example: X = the number of telephone calls in an hour.

$$PDF: \quad p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \cdots; \lambda > 0$$

Examples:

1. The manufacturer of the disk drives in one of the well-known brands of microcomputers expects 2% of the disk drives to malfunction during the microccomputer's warranty period. Calculate the probability that in a sample of 100 disk drives, that not more than three will malfunction.

No. of disk drives malfunctioning k	Binomial $\binom{100}{k}.02^k.98^{100-k}$	Poisson Approximation $e^{-2}2^k/k!$
0	0.13262	0.13534
1	0.27065	0.27067
2	0.27341	0.27067
3	0.18228	0.18045
Total	0.85890	0.85713

2. The average rate of job submissions in a busy computer centre is 4 per minute. If it can be assumed that the number of submissions per minute interval is Poisson distributed, calculate the probability that
   (a) at least 2 jobs will be submitted in any minute.
   (b) no job will be submitted in any minute.
   (c) no more than one job will be submitted in any one-minute interval.

Poisson as an approximation to the binomial when n is large p is small
Recall:

- mean of binomial $= np$
- mean of Poisson $= \lambda$

PDF of Binomial

$$\begin{aligned} P(x) &= \binom{n}{x} p^x (1-p)^{n-x}; \quad p = \frac{\lambda}{n} \\ &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \frac{(1-\frac{\lambda}{n})^n}{\left(1-\frac{\lambda}{n}\right)^x} \to \frac{\lambda^x}{x!} e^{-\lambda} \end{aligned}$$

Now

$$\begin{aligned} (1-\frac{\lambda}{n})^n &\to e^{-\lambda} \text{ as } n \to \infty \\ (1-\frac{\lambda}{n})^x &\to 1 \text{ as } n \to \infty \\ \frac{n!}{(n-x)!n^x} &= \frac{n(n-1)\ldots(n-(x-1))}{n^x} \\ &= 1(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{x-1}{n}) \to 1 \end{aligned}$$

Events Distributed Independently in Time

1. Probability that an event occurs in a very small time interval, of length $\triangle t$, is $\lambda\triangle t$. i.e. $P_1(\triangle t) = \lambda\triangle t$.

2. Probability of more than one occurrence in the small time interval is negligible. $P_{>1}(\triangle t) = 0. \Longrightarrow P_0(\triangle t) = 1 - \lambda\triangle t$.

3. Occurrences are independent in different time intervals

X = number of events occurring in a time interval of length t
We will show that the probability of x events in t

$$P_x(t) = e^{-\lambda t}(\lambda t)^x/x!$$

Note

$$\begin{aligned} P_0(t + \triangle t) &= P_0(t)P_0(\triangle t) = P_0(t)(1 - \lambda\triangle t) \\ P_x(t + \triangle t) &= P_x(t)P_0(\triangle t) + P_{x-1}(t)P_1(\triangle t) \\ &= P_x(t)(1 - \lambda\triangle t) + P_{x-1}(t)\lambda\triangle t,\ \ x = 1, 2, 3, \ldots \end{aligned}$$

First show

$$P_0(t) = e^{-\lambda t}$$

Proof:

$$P_0(t + \triangle t) = P_0(t)P_0(\triangle t) = P_0(t)(1 - \lambda\triangle t)$$

Thus

$$P_0(t + \triangle t) - P_0(t) = -\lambda\triangle t P_0(t)$$

and

$$\frac{P_0(t + \triangle t) - P_0(t)}{\triangle t} = -\lambda P_0(t)$$

Let $\triangle t \to 0$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$

$$\int \frac{1}{P_0(t)} dP_0(t) = \int -\lambda dt$$

$$\log P_0(t) = -\lambda t + c \implies P_0(t) = e^{-\lambda t + c}$$

Now $c = 0$, since $P_0(0) = 1$ (the probability of 0 occurrences in a zero time interval)
So

$$P_0(t) = e^{-\lambda t}$$

Next show

$$P_1(t) = e^{-\lambda t}\lambda t$$

Proof:

Look at

$$\begin{aligned} P_1(t + \triangle t) &= P_1(t)P_0(\triangle t) + P_0(t)P_1(\triangle t) \\ &= P_1(t)(1 - \lambda\triangle t) + e^{-\lambda t}\lambda\triangle t \end{aligned}$$

So

$$\frac{P_1(t + \triangle t) - P_1(t)}{\triangle t} = -\lambda P_1(t) + e^{-\lambda t}\lambda$$

Now as $\triangle t \to 0$ we have

$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda e^{-\lambda t}$$

Multiply both sides by the integrating factor $e^{\lambda t}$

$$\begin{aligned} e^{\lambda t}\left[\frac{dP_1(t)}{dt} + \lambda P_1(t)\right] &= \lambda \\ \Longrightarrow \qquad \frac{de^{\lambda t}P_1(t)}{dt} &= \lambda \\ \Longrightarrow \qquad e^{\lambda t}P_1(t) &= \lambda t + c \text{ (check } c = 0) \end{aligned}$$

Thus

$$P_1(t) = e^{-\lambda t}\lambda t$$

16 SAMPLING INSPECTION

A method of deciding if a batch of components is to be accepted or rejected, on the basis of the number of defectives found in a random sample taken from the batch

Example:
Samples of size 10 are taken from a series of large batches in an industrial process. The batch is accepted if the sample contains no defectives, otherwise it is rejected

This is called a **Single Sampling Scheme** since the decision on each batch is based on an inspection of a single sample

The probability (P) of accepting batches can be calculated for various proportions defective.

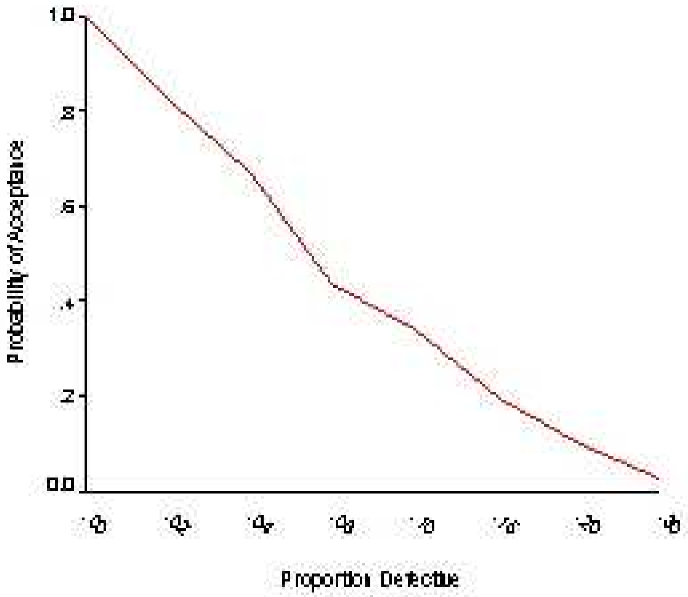
Acceptance Rule $P = P(0\ defectives) = q^{10}$

p	.00	.02	.04	.08	.10	.15	.20	.30
P	1.00	.817	.667	.434	.348	.197	.108	.028

A plot of P against p is called

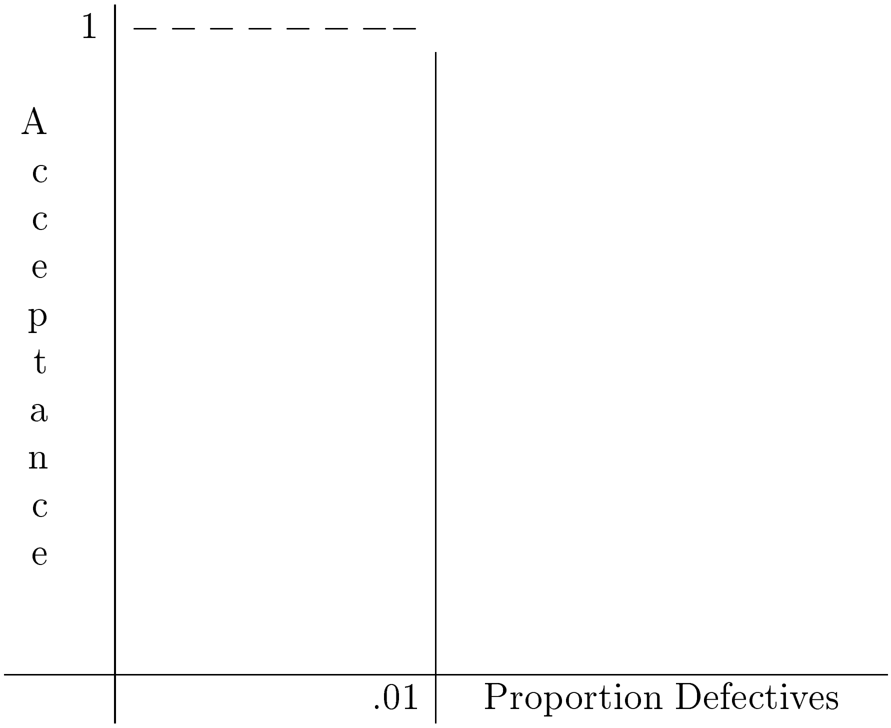
the Operating Characteristic Curve

Figure 15: The OC Curve



Ideal OC curve

Suppose batches of up to but not exceeding 1% are acceptable, the OC curve should look like this



This can only be achieved by 100% inspection.

Design of Sampling Scheme

A sampling scheme consists of

1. sample size
2. allowable number of defectives

It is chosen usually after specifying the risks of the consumer and producer

Risks:

- **Consumer's Risk:**
  Accept lots which should be rejected

- **Producer's Risk:**
  Reject lots which should be accepted

The probability of acceptance (P) depends on:

1. unknown proportion of defectives in the batch

2. sampling scheme.

Small Batches

Example: A certain item is packed twenty to a box, and the following single sampling plan is used: a sample of 2 items is drawn and the box is accepted if both items are good, otherwise the box is rejected.
Construct the OC curve
Solution:

$$N = 20 \quad n = 2$$

Sampling is done without replacement $\longrightarrow$ use hypergeometric distribution.
$p =$ proportion defective
$q = 1 - p =$ proportion non - defective
Hence batch looks like

Np	Nq

Generally with c= allowable number of defectives:

$$P(accept) = P = \sum_{x=0}^{c} \binom{Np}{x} \binom{Nq}{n-x} / \binom{N}{n}$$

In the above example:

$$P = \binom{20p}{0} \binom{20q}{2} / \binom{20}{2}$$

Calculate P for $p = .1\ .2\ .3\ .4\ .5$ and plot P against p

Average Outgoing Quality (AOQ) of Rectifying Scheme

Defn: Rectifying Scheme:

A rectifying scheme is where lots are inspected and the defectives are replaced.

Defn: AOQ = average quality of outgoing products including

1. accepted lots
2. rejected lots after the rejected lots have been 100% inspected and replaced by non-defectives

Example:

$$\begin{aligned} \text{Lot size} &= N \\ \text{Sample size} &= n \end{aligned}$$

This lot breaks down as follows :

1. sample of size n
2. $N - n$ items which are either
   (a) inspected 100%
   (b) accepted without further inspection

In (1) all defectives are replaced
In (2a) all defectives are replaced
But in (2b) the defectives are not replaced

Average Outgoing Quality (AOQ)

Suppose p = proportion defective
then $p(N - n)$ = number defective in the lot after a sample of size n has been drawn.

If P = probability of accepting lots
Then, **in the long run**, **on the average**, the expected number of defectives in the lot is

$$Pp(N - n)\text{defective items}$$

The AOQ is the average number of defectives expressed as a proportion of the lot size

$$\begin{aligned} AOQ &= \frac{\text{Expected no. of defectives per batch}}{\text{Batch Size}} \\ &= \frac{Pp(N-n)}{N} \\ &\approx Pp \text{ when } N \text{ is large} \end{aligned}$$

Example:

Lot Size: $N = 10,000$
Sampling Plan:
Allow up to two defectives in a sample of size 100.

OC Table

p	0	.01	.02	.03	.04	.05	.06	.07
P	1	.92	.67	.42	.24	.13	.06	.03

$$\begin{aligned} P[\text{acceptance}] &= P(0) + P(1) + P(2) \\ &= e^{-\lambda} + e^{-\lambda}\lambda + e^{-\lambda}\frac{\lambda^2}{2!} \text{ where} \lambda = 100p \end{aligned}$$

Figure 16: OC Curve

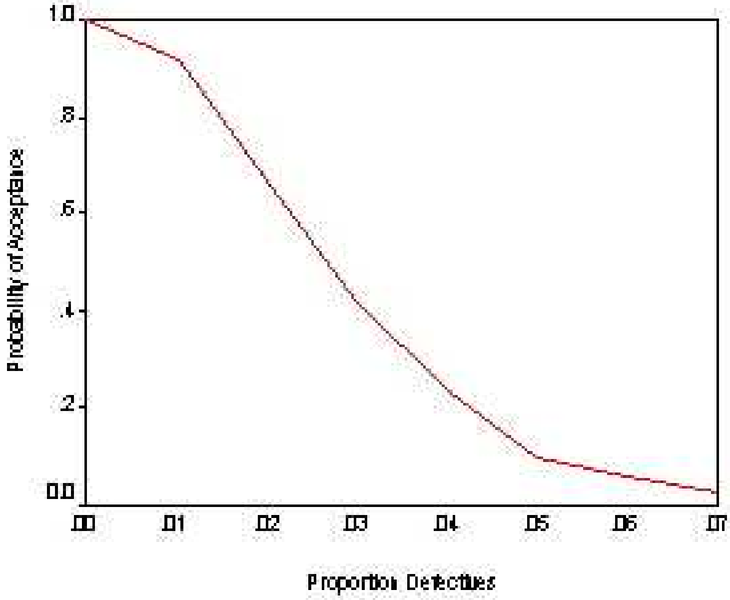
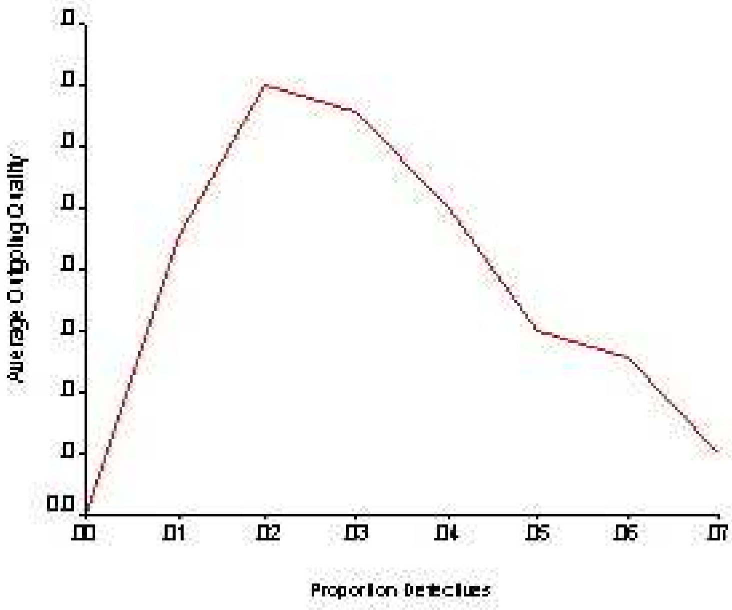


Figure 17: AOQ



Example:

Lot Size: $N = 10,000$
Sampling Plan:
Allow up to two defectives in a sample of size 100.

Average Outgoing Quality

p	0	.01	.02	.03	.04	.05	.06	.07
Pp	0	.009	.014	.013	.010	.006	.005	.002

$$\textbf{AOQL} = \text{Average Outgoing Quality Limit}$$
$$= \text{Worst Possible Quality}$$

17 DOUBLE SAMPLING SCHEMES

First sample can lead to one of three decisions

1. accept lot
2. take a second sample
3. reject lot

Example 1

Stage 1: Take a sample of size 10

1. if 0 defectives occurs - accept lot
2. if 1,2 or 3 defectives occur - take a second sample
3. if 4 or more defectives occur - reject lot

Stage 2: If 1, 2, 3 defectives occur in first sample take a second sample of size 10 and reject lot if the total number of defectives in the combined sample is 4 or more

Example 1

1. If 0 defectives occur - accept lot
2. If 1,2 or 3 defectives occur - take a second sample
3. If 4 or more defectives occur - reject lot

Stage 2:

1. If ≥ 4 in combined sample reject.
2. If < 4 accept.

The OC Curve

$$\begin{aligned} \text{Let E} &= \text{ event of accepting batch} \\ A &= \text{ accept batch at first stage} \\ B &= \text{accept batch at second stage} \end{aligned}$$

Then

$$P(E) = P(A \cup B) = P(A) + P(B)$$

$$\begin{aligned} P(A) &= P(\text{0 defect}) = (1-p)^{10} \\ P(B) &= P(\text{1 defect in 1st sample})P(\leq \text{2 defects in 2nd sample}) \\ &+ P(\text{2 defects in 1st sample})P(\leq \text{1 defect in 2nd sample}) \\ &+ P(\text{3 defects in 1st sample})P(\text{0 defect in 2nd sample}) \end{aligned}$$

Calculate $P(A) + P(B)$ for various values of p to obtain the OC curve

The average sample size

$$\overline{n} = 10 + 10 * P(\text{taking a 2nd sample})$$

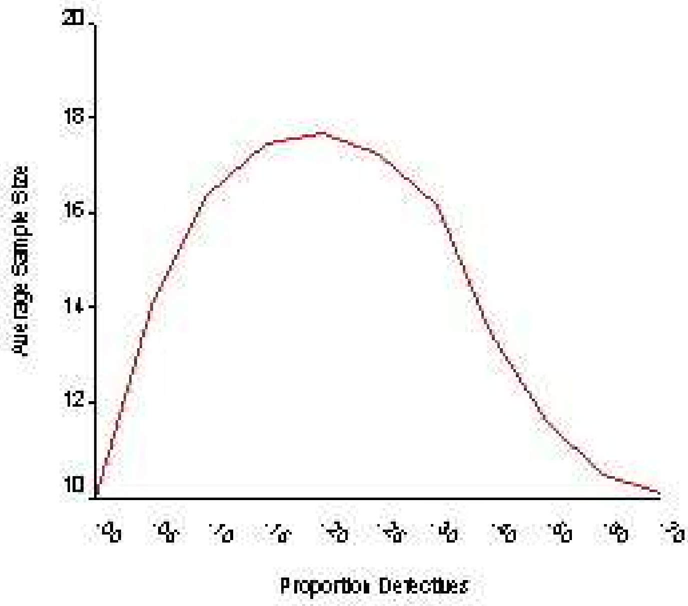
Now:

$$\begin{aligned} & P(\text{taking a 2nd sample}) \\ = \; & P(1, 2 \text{ or } 3 \text{ defectives in first stage}) \\ = \; & \binom{10}{1}(1-p)^9 p + \binom{10}{2}(1-p)^8 p^2 + \binom{10}{3}(1-p)^7 p^3 \end{aligned}$$

$\overline{n}$ is calculated for various values of p

p	.00	.05	.10	.15	.20	.25	.30
$\overline{n}$	10.0	14.0	16.4	17.5	17.7	17.2	16.2

Figure 18: Average Sample Size



18 CONTINUOUS DISTRIBUTIONS

The PDF of a continuous distribution

1. the total area under the PDF is 1
2. the area under the PDF between two points a and b is the probability that the random variable lies between a and b
3. PDF is positive or zero - it is zero in any range where the random variable never falls

Notation:

- **Probability Density Function (PDF)**:
  The PDF of a continuous variable X is usually denoted by $f(x)$

- **Cumulative Distribution Function (CDF)**
  As in the discrete case, the CDF is defined as the probability that the random variable X is less than or equal to some specified x.

  The CDF of a continuous variable X is usually denoted by $F(x)$ and defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(X)dX$$

The Exponential Distribution

- **The PDF:**

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t}, \ t > 0 \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$

- **The CDF:**

$$
\begin{aligned}
P(T \leq t) = F(t) &= \int_0^t \lambda e^{-\lambda T} dT \\
&= \left[-e^{-\lambda T} \right]_{T=0}^{t} \\
&= -e^{-\lambda t} + 1
\end{aligned}
$$

i.e.

$$
\begin{aligned}
F(t) &= 1 - e^{-\lambda t}, \quad t > 0 \\
&= 0 \ \ \text{otherwise.}
\end{aligned}
$$

Generally the exponential distribution describes waiting time between Poisson occurrences

Proof:
Let T = time that elapses after a Poisson event.

$P(T > t)$ = probability that no event occurred in the time interval of length t.

The probability that no Poisson event occurred in the time interval $[0, t]$:

$$P(0, t) = e^{-\lambda t}.$$

where λ is the average Poisson occurrence rate in a unit time interval.
So:

$$P(T > t) = e^{-\lambda t},$$

Hence the CDF is:

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t}$$

and, working backwards, the PDF is

$$f(t) = F'(t) = \lambda e^{-\lambda t}$$

The Exponential Distribution:

Examples:

1. Time between telephone calls

2. Time between machine breakdowns

3. Time between successive job arrivals at a computing centre

Example
Accidents occur with a Poisson distribution at an average of 4 per week. i.e. $\lambda = 4$

1. Calculate the probability of more than 5 accidents in any one week

2. What is the probability that at least two weeks will elapse between accident?

Solution

1. Poisson:
$$P(X > 5) = 1 - P(X \leq 5)$$

2. Exponential:
$$P(\text{Time between occurrences} > 2)$$
$$= \int_2^\infty \lambda e^{-\lambda T} dT$$
$$= \left[-e^{-\lambda T}\right]_{T=2}^{\infty} = e^{-8} = .00034$$

The Markov Property of Exponential

Examples:

1. The distribution of the remaining life does not depend on how long the component has been operating.

   i.e.
   the component does not ‘age’ - its breakdown is a result of some sudden failure not a gradual deterioration

2. Time between telephone calls

   Waiting time for a call is independent of how long we have been waiting

Applications of the Exponential

Empirical investigation has shown that the following experiments have exponential distributions

1. length of telephone calls

2. service time at a server in the queueing network.

3. lifetime of components

The Markov property
Show:

$$P(T \leq x + t | T > t) = P(T \leq x)$$

Proof:

$$E_1 = T \leq x + t, \text{and} \quad E_2 = T > t$$

Then:

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

Now

$$\begin{aligned} P(E_1 \cap E_2) &= P(t < T \leq x + t) \\ &= \int_t^{x+t} \lambda e^{-\lambda T} dT \\ &= e^{-\lambda t}[1 - e^{-\lambda x}] \end{aligned}$$

and

$$P(E_2) = \int_t^{\infty} \lambda e^{-\lambda T} dT = e^{-\lambda t}$$

thus

$$P(E_1|E_2) = \frac{e^{-\lambda t}[1 - e^{-\lambda x}]}{e^{-\lambda t}} = 1 - e^{-\lambda x}$$

now

$$1 - e^{-\lambda x} = F(x) = P(T \leq x)$$

Examples

1. Calls arrive at an average rate of 12 per hour. Find the probability that a call will occur in the next 5 minutes given that you have already waited 10 minutes for a call i.e. Find $P(T \leq 15|T > 10)$

   From the Markov property

   $$P(T \leq 15|T > 10) = P(T \leq 5)$$

   So:

   $$P(T \leq 5) = 1 - e^{-5\lambda}$$

   The average rate of telephone calls is $\lambda = \cdot 2$ in a minute, then

   $$\begin{aligned} P(T \leq 5) &= 1 - e^{-(5)(\cdot 2)} \\ &= 1 - e^{-1} \end{aligned}$$

2. The average rate of job submissions in a busy computer centre is 4 per minute. If it can be assumed that the number of submissions per minute interval is Poisson distributed, calculate the probability that:

   (a) at least 15 seconds will elapse between any two jobs.

   (b) less than 1 minutes will elapse between jobs.

   (c) If no jobs have arrived in the last 30 seconds, what is the probability that a job will arrive in the next 30 seconds.

19 EXPECTATION OF CONTINUOUS VARIABLES

Recall, when X is discrete

$$E(X) = \sum_x xp(x)$$

Similarly, when X is continuous

$$E(X) = \int_x xf(x)dx$$

It can be easily be shown

1. E(X+c) = E(X) + c

$$\begin{aligned} E(X+c) &= \int_x (x+c)f(x)dx \\ &= \int_x xf(x)dx + c\int_x f(x)dx \\ &= E(X) + c. \end{aligned}$$

2. E(cX)= cE(X)

$$\begin{aligned} E(cX) &= \int_x cxf(x)dx \\ &= c\int_x xf(x)dx \\ &= cE(X). \end{aligned}$$

3. V(X+c) = V(X)

$$\begin{aligned} V(X+c) &= E[X+c-E(X+c)]^2 \\ &= E[X-E(X)]^2 \\ &= V(X) \end{aligned}$$

4. $V(cX) = c^2V(X)$

$$\begin{aligned} V(cX) &= E[cX-E(cX)]^2 \\ &= E[cX-cE(X)]^2 \\ &= E[c(X-E(X))]^2 \\ &= c^2E[X-E(X)]^2 \\ &= c^2V(X) \end{aligned}$$

The exponential distribution

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t},\ t > 0 \\ &= 0 \ \ \text{otherwise.} \end{aligned}$$

Show that

$$E(T) = \int_0^{\lambda} e^{-\lambda T} dT = \frac{1}{\lambda}$$

Example:
average number of telephone calls received in an hour = 4

$\Longrightarrow$ average waiting time between telephone calls is

$$\frac{1}{4} \text{ hours} \quad \text{or 15 minutes}$$

20 THE NORMAL DISTRIBUTION

The normal distribution measures continuous variables such as heights, weights.
The PDF is

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

It can be shown that

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

X is said to be normally distributed with a mean of μ and a variance of σ^2.

i.e.

$$X \sim N(\mu, \sigma)$$

THE STANDARD NORMAL DISTRIBUTION

$$f(Z) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-Z^2}{2}\right) \qquad -\infty < Z < \infty$$

It this case

$$\begin{aligned} E(Z) &= 0 \\ V(Z) &= 1 \end{aligned}$$

Z is said to be normally distributed with a mean of 0 and a variance of 1.

i.e.

$$Z \sim N(0, 1)$$

Clearly if

$$X \sim N(\mu, \sigma),$$

then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

The PDF and the CDF of the standard normal variable are tabulated.

Reading Standard Normal Tables

1. $P(Z > 0)$

2. $P(Z > 1)$

3. $P(Z > -1)$

4. $P(-1 < Z < 1)$

5. $P(-1.96 < Z < 1.96)$

6. $P(Z < 2.58)$

Choose k so that

1. $P(-k < Z < k) = .95$

2. $P(-k < Z < k) = .99$

3. $P(-k < Z < k) = .68$

21 APPLICATIONS OF NORMAL DISTRIBUTION

Examples:
Exam results are normally distributed the mean $\mu = 46$ and the standard deviation $\sigma = 4$ what percentage of students obtain a mark

1. larger than 46?
2. larger than 50?
3. larger than 40?
4. less than 38?
5. less than 49?
6. between 45 and 49?
7. between 50 and 54?
8. larger than 56 or less than 40?
9. within 1.5 standard deviations from the mean?
10. outside of 2.3 standard deviation from the mean?

Example:
Assume the scores on an aptitude are normally distributed with mean $\mu = 500$ and standard deviation $\sigma = 100$.

1. What is the top 5% cut off point?

2. What is the middle 40%?

3. If 1000 new students are to take the exam, predict the number who will score more than 65%.

Conditional Probabilities
Example:
The lifetime of a transistor is normally distributed with mean $\mu = 200$ and a standard deviation $\sigma = 9$.

What is the probability that a transistor

1. will last more than 220 hours?

2. will last between 210 and 220 hours?

3. which lasts 210 hours will last at least 220 hours?

4. which lasts 210 hours will fail within 10 hours. (i.e. will fail before 220 hours)

22 INEQUALITIES

Markov Inequality
Let X be a random variable with expected value $E(X)$ known and such that $P(X < 0) = 0$ then, for any $k > 0$,

$$P(X \geq k) \leq E(X)/k$$

Example:

An online computer system is proposed for which it is estimated that the mean response time is:

$$E(T) = 10 \text{ seconds}$$

Estimate the probability that the response time will be more than 20 seconds.

Proof of Markov's Inequality

Discrete Case:

$$
\begin{aligned}
E(X) &= \sum_x xp(x) \\
&= \sum_{x \geq k} xp(x) + \sum_{x<k} xp(x) \\
&\geq \sum_{x \geq k} xp(x) \text{ since } x \text{ is non negative} \\
&\geq \sum_{x \geq k} kp(x) \\
&\geq k \sum_{x \geq k} p(x) \\
&= kP(x \geq k)
\end{aligned}
$$

i.e.

$$E(X) \geq kP(x \geq k)$$

i.e.

$$P(X \geq k) \leq E(X)/k$$

Proof of Markov's Inequality

Continuous Case

$$
\begin{aligned}
E(X) &= \int_x xf(x)d(x) \\
&= \int_{x\geq k} xf(x)dx + \int_{x<k} xf(x)dx \\
&\geq \int_{x\geq k} xf(x)dx \\
&\geq \int_{x\geq k} kf(x)dx \\
&= kP(x \geq k)
\end{aligned}
$$

i.e.

$$P(X \geq k) \leq E(X)/k$$

Tchebychev's Inequality

If X is a random variable with mean μ and variance σ then for any k

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Example

An on-line computer system is proposed for which it is estimated that the mean response time

$$\mu = E(t) = 10 \text{ seconds}$$

and the variance

$$\sigma^2 = 4 \ \text{ seconds.}$$

Estimate the probability that the response will be between 4 and 16 seconds.

Proof of Tchebychev's Inequality

Discrete case

$$\begin{aligned}
\sigma^2 &= E(X-\mu)^2 \\
&= \sum_x (x-\mu)^2 p(x) \\
&= \sum_{|x-\mu|\geq k\sigma} (x-\mu)^2 p(x) + \sum_{|x-\mu|<k\sigma} (x-\mu)^2 p(x) \\
&\geq \sum_{|x-\mu|\geq k\sigma} (x-\mu)^2 p(x) \\
&\geq \sum_{|x-\mu|\geq k\sigma} (k\sigma)^2 p(x) \\
&\geq k^2\sigma^2 \sum_{|x-\mu|\geq k\sigma} p(x) \\
&= k^2\sigma^2 P(|x-\mu| \geq k\sigma)
\end{aligned}$$

i.e.

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Proof of Tchebychev's Inequality

Continuous Case

$$\begin{aligned}
\sigma^2 &= E(X-\mu)^2 \\
&= \int_x (x-\mu)^2 f(x)dx \\
&= \int_{|x-\mu|\geq k\sigma} (x-\mu)^2 f(x)dx + \int_{|x-\mu|<k\sigma} (x-\mu)^2 f(x)dx \\
&\geq \int_{|x-\mu|\geq k\sigma} (x-\mu)^2 f(x)dx \\
&\geq \int_{|x-\mu|\geq k\sigma} (k\sigma)^2 f(x)dx \\
&\geq k^2\sigma^2 \int_{|x-\mu|\geq k\sigma} f(x)dx \\
&= k^2\sigma^2 P(|X-\mu| \geq k\sigma)
\end{aligned}$$

i.e

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Examples

1. A model of an on-line computer system gives a mean time to retrieve a record from a direct access storage system device of 200 milliseconds with a standard deviation of 58 milliseconds. The design criterion requires that at least 90% of all retrieval times must not differ from the mean by more that 75 milliseconds.

   (a) Use Tchebychev's inequality to establish whether the design criterion is satisfied.

   (b) Would the design criterion be satisfied if it were known that the retrieval time is normally distributed with a mean of 200 milliseconds and a standard deviation of 58 milliseconds?

2. A model of an on-line computer system gives a mean time to retrieve a record from a direct access storage system device of 100 milliseconds. If a record is retrieved, calculate the probability that its retrieval time will be:

   (a) at least 150 milliseconds;

   (b) between 150 and 200 milliseconds.