BHARATHIDASAN UNIVERSITY

(Re-accredited with 'A' Grade by NAAC)

CENTRE FOR DISTANCE EDUCATION

PALKALAIPERUR, TIRUCHIRAPPALLI – 24

M.B.A

CORE COURSE VII

II – Semester

OPERATIONS RESEARCH

Chairman:

Dr. V.M Muthukumar
Vice-Chancellor
Bharathidasan University
Tiruchirappalli – 620 024.

Co-Ordinator:

Dr. R.Babu Rajendren
Registrar i/c
Centre for Distance Education
Bharathidasan University
Tiruchirappalli – 620 024.

Course Director:

Dr.V.Vinod Kumar
Director i/c
Centre for Distance Education
Bharathidasan University
Tiruchirappalli – 620 024.

The Syllabus Revised from 2017-18 onwards

Lesson Writer:

Dr.S.A.Sirajudeen
Associate Professor & Head,
PG&Research
Dept.of Management Studies,
J.J College of Arts & Science (A)
Pudukottai – 622 422.

CORE COURSE VII
OPERATION RESEARCH

Objectives:

The objectives of the course is to acquaint the student with the applications of Operations Research to business and industry and help them to grasp the significance of analytical techniques in decision making. Students will be tested on the application of Operations Research to business related problems.

Unit- I

Introduction to Operations Research, scope, phases-merits and limitations – concept of optimization.Theory of simplex methods to solve canonical and general LPP, Primal – Dual problem and its properties, dual simplex method, sensitivity analysis.Concept of goal programming.

Unit- II

Transportation problem by Vogel's approximation method; assignment problem , linear Programming complete enumeration method.

Unit- III

Network analysis – drawing of arrow diagram – critical path method – calculation of critical path duration , total , free and independent floats , PERT problems; Inventory Theory , Deterministic models – purchase problem without and with shortages , with price breaks , production problem without shortages.

Unit- IV

Decision under risk – expected money value criterion – decision trees – decision under uncertainty – minimax criterion; Theory of Games – pure and mixed strategies, Principles of dominance , graphical methods , simplex methods.

Unit- V

Queuing theory – M/M/1/FIFO/oc model; Markovian chain, Simulation :- Monte Carlo Method.

UNIT-I

INTRODUCTION TO OPERATION RESEARCH

This chapter provides an overall view of the subject of operation research. It covers some general ideas on the subject , thus providing a perspective. The remaining chapters deal Wwith specific ideas and specific methods of solving OR problems.

1.1 DEVELOPMENT OF OPERATIONS RESEARCH

(i)Pre-World War II: No science has ever been born on a specific day. Operations Research is no exception. Its roots are as old as science and society. Though the roots of OR extend to even early 1800s, it was in 1885 when Ferderick W. Taylor emphasised the application of scientific analysis to methods of production, that the real start took place. Taylor conducted experiments in connection with the simple shovel. His aim was to find that weight load of ore moved by shovel which would result in maximum of ore moved with minimum of fatigue. After many experiments with varying weights, he obtained the optimum weight load, which though much lighter than that commonly used, provided maximum movement of ore during a day.

Another man of early scientific management era was Henry L. Gantt. Most job-scheduling methods at that time ere rather haphazard. A job, for instance, may be processed on a machine without trouble but then wait for days for acceptance by the next machine. Gantt mapped each job from machine to machine, minimizing every delay. Now, with the Gantt procedure it is possible to plan machine loadings months in advance and still quote delivery dates accurately.

The well-known economic lot size model is attributed to F.W. Harris, who published his work on the area of inventory control in 1915.

OPERATIONS RESEARCH IN INDIA

In India, operations research came into existence with the opening of an OR unit in 1949 at the Regional Research Laboratory in Hyderabad. An OR unit under professor P.C. Mahalonobis was established in 1953 in the Indian Statistical Institute, Kolkata to apply OR methods in national planning and survey. Operations Research society of india was formed in 1957 and its first conference was held in Delhi in 1959. It was felt that there existed the need of producing well-trained operations researchers eho could tackle practical problems. It was also decided to bring out a journal on operations research, the first volume of which came out in 1963 with the name ‘Opsearch’. Some other indian journals promoting the cause of operations research are ‘Industrial Engineering and Management’ , ‘Materials Management Journal of India’, ‘Defence Science Journal’, ‘Journal of the Indian Society of Statistics and Operations Research’ (ISSOR), ‘Pure and Applied Mathematika Sciences’ (PAMS), etc.

Professor Mahalonobis made the first important application of OR in india in preparing the draft of the second five year plan. The draft plan frame is still the most scientifically formulated plan bearing programmes of massive economic development of india. It was estimated that india could become self-sufficient in food merely by reducing its wastage by 15%.

For academic studies, the first M.Sc. Course on OR was started by Delhi University in 1963. At the same time, Institutes of Management at Kolkata and Ahmedabad introduced OR in their MBA courses. At present, this subject has been introduced in almost all Institutes and Universities for the students of mathematics, statistics, commerce, economics, management and engineering. Realising the importance of OR in accounts and administrations, the government has introduced this subjects for CA, ICWA and IAS examinations.

1.2 DEFINITIONS OF OPERATIONS RESEARCH

Operation research, rather simply defined, is the research of operations. An operation may be called a set of acts required for the achievement of a desired outcome.

Many definitions of OR have been suggested from time to time. On the other hand are put forward a number of arguments as to why it cannot be defined. Perhaps the subject has too wide scope of applications to be defined in a precise manner. Also it is not easy to define operations research precisely as it is not a science representing any well-defined social, biological or physical phenomenon. Some of the different definitions suggested are:

(1) OR is a *scientific method* of providing executive departments with a *quantitative* basis for decisions regarding the operations under their control.

--Morse & Kimball

(2) OR, in the most general sense, can be characterised as the application of *scientificmethods, tools and techniques* to problems involving the *operations of systems* so as to provide those in control of the operations with optimum solutions to the problems.

--Churchman Ackoff, Arnoff

(3) Operations research is applied *decision theory*. It uses any *specific, mathematicalor logical means* to attempt to cope with the problems that confront the executive when he tries to achieve a thorough going rationality in dealing with his decision problems.

--Miller and Starr

(4) Operations research is a *scientific* approach to problem solving for executive management.

--H.M. Wagner

(5) Operations research is the art of giving *bad answers* to problems, to which, otherwise, *worse answers* are given. -- Thomas L. Saaty

(6) Operations research is the *art of winning wars without actually fighting them*.

--Auther Clark

(7) Operations research is an aid for the executive in making his decisions by providing him with the needed *quantitative information* based on the *scientific method of analysis.*

--C.Kittel

(8) Operations research is the *systematic, method-oriented*study of the basic structure, characteristics, functions and relationships of an organization to provide the executive with a *sound, scientific and quantitative basis* for decision-making.

--E.L. Arnoff& M.J. Netzorg

(9) Operations research is the application of scientific methods to problems arising from operations involving *integrated systems of men, machines and materials*. It normally utilizes the knowledge and skill of an *interdisciplinary research team* to provide the managers of such systems with *optimum operating solutions*.

--Fabrycky&torgerson

(10) Operations research is an experimental and applied science devoted to observing, understanding and predicting the behaviour of purposeful man-machine systems, and operations research workers are actively engaged in applying this knowledge to practical problems in business, government and society

--Operations Research Society of America

1.3 CHARACTERISTICS OF OPERATIONS RESEARCH

The various definitions of operations research presented in section 1.2 bring out the essential characteristics of operations research. They are:

(i) its system (or executive) orientation,

(ii) The use of interdisciplinary teams,

(iii) Application of scientific method,

(iv) Uncovering of new problems,

(v) Improvement in the quality of decisions,

(vi) Use of computer,

(vii) Quantitative solutions, and

(viii) Human factors.

Let us consider each of these in some detail.

1.3-1 SYSTEM (OR EXECUTIVE) ORIENTATION OF OR

One of the most important characteristics of OR study is its concern with problems as a whole or its system orientation. This means that an activity by any part of an organization has some effect on the activity of every other part. The optimum operation of one part of a system may not be the optimum operation for some other part. Therefore, to evaluate any decision, one must identify all possible interactions and determine their impact on the organization as a whole.

Many problems that appear simple on the surface may not be really so. Take, for example, the inventory policy of an organization, already considered in section 1.1. production department is interested in long, uninterrupted production runs since they reduce the set-up and clean-up costs. To solve the problem with this viewpoint is simple. However, these long runs will result in large raw material, in-process and finished product inventories in relatively few product lines. This will result in bitter conflict with finance, marketing and personnel departments. As already discussed finance department wants to have the minimum possible inventory; marketing department, a large but diversified inventory, while personnel department wants continuous production during slack periods also, resulting in large inventories.

In view of the above difficulties, it is necessary that the problem be analysed with painstaking care and all parts of the organization affected be thoroughly examined.

1.3-2 THE USE OF INTERDISCIPLINARY TEAMS

The second characteristic of OR study is that it is performed by a team of scientist whose individual members have been drawn from different scientific and engineering disciplines. For example, one may find a mathematician, statistician, physicist, psychologist, economist, and an engineer working together on a OR problem.

It has been recognised beyond doubt that people from different disciplines can produce more unique solutions with greater probability of success, than could be expected from the same number of persons from a single discipline. For example, when confronted with the problem of increasing production in a plant, the personnel psychologist will try to select better workers or improve their training; the mechanical engineers will try to improve the machines; the industrial engineer will try to simplify the operations or offer incentives; while the systems analyst will try to improve the flow of information from the plant. Thus the OR team can look at the problem from many different angles in order to determine which one (or which combination) of approaches is the best.

Another reason for the existence of OR team is that knowledge is increasing at a very fast rate. No single person can collect all the useful scientific information from all disciplines. Different members of the OR team bring the latest scientific know-how in different disciplines to analyse the problem andhelp in providing better results.

1.3-3 APPLICATION OF SCIENTIFIC METHOD

The third distinguishing feature of OR is the use of scientific method to solve the problem under study. Most scientific research, such as chemistry and physics can be carried out well in the laboratories, under controlled conditions, without much interference from the outside world. However, the same is not true in the systems under study by OR team. For example, no company can risk its failure in order to conduct a successful experiment. Though, experimentation on sub-systems is sometimes resorted to, by and large, a research approach that does not involve experimentation on the total system is preferred.

An operations research worker is in the same position as the astronomer, since the latter can observe the system that he studies, but cannot manipulate it. Therefore, he constructs *representation of systems and its operations (models)* on which he conducts his research. An OR worker also does the same. The construction of a model is described in section 1.12-2 and the reader may refer it for further details.

1.3-4 UNCOVERING OF NEW PROBLEMS

The fourth characteristic of operation research, which is often overlooked, is that solution of an OR problem may uncover a number of new problems. Of course, all these uncovered problems need not be solved at the same time. However, in order to derive maximum benefit, each one of them must be solved. It must be remembered that OR is not effectively used if it is restricted to one-shot problems only. In order to derive full benefits, continuity of research must be maintained. Of course, the results of OR study pertaining to a particular problem need not wait until all the connected problems are solved.

1.3-5 IMPROVEMENT IN THE QUALITY OF DECISIONS

OR gives bad answers to problems, to which, otherwise, worse answers are given. It implies that by applying its scientific approach, it can only improve the quality of solution but it may not be able to give perfect solution.

1.3-6 USE OF COMPUTER

Another characteristic of OR is that it often requires a computer to solve the complex mathematical model or to manipulate a large amount of data or to perform a large number of computations that are involved.

1.3-7 QUANTITATIVE SOLUTIONS

OR approach provides the management with a quantitative basis for decision-making. For example, it will give answer like, "the cost to the company, if decision A is taken is X; if decision B is taken is Y, etc."

1.3-8 HUMAN FACTORS

In deriving quantitative solutions we do not consider human factors, which doubtlessly play a great role in the problems posed. Definitely an OR study is incomplete without a study of human factors.

1.4 NECESSITY OF OPERATIONS RESEARCH IN INDUSTRY

After having studied as to what is operations research, we shall now try to answer as to why study OR or what is its importance or why its need has been felt by the industry.

As already pointed out, science of OR came into existence in connection with the war operations, to decide the strategy by which enemy could be harmed to the maximum possible extent with the help of the available warfare. War situation required reliable decision-making. But its need has been equally felt by the industry due to the following reasons:

(a)***Complexity****:* In a big industry, the number of factors influencing a decision have increased. Situation has become big and complex because these factors interact with each other in complicated fashion. There is, thus, great uncertainty about the outcome of interaction of factors like technological, environmental, competitive, etc.For instance, consider a factory production schedule which has to take into account

(i) customer demand,

(ii) requirements of raw materials,

(iii) equipment capacity and possibility of equipment failure, and

(iv) restrictions on manufacturing processes.

Evidently, it is not easy to prepare a schedule which is both economical and realistic. This needs mathematical models, which, in addition to optimization, help to analyse the complex situation. With such models, complex problems can be split up into simpler parts, each part can be analysed separately and then the results can be synthesized to give insights into the problem.

(b)***Scattered responsibility and authority***: In a big industry, responsibility and authority of decision-making is scattered throughout the organization and thus the organization, if it is not conscious, may be following inconsistent goals. *Mathematical quantification* of OR overcomes this difficulty also to a great extent.

(c)*Uncertainty*: There is a great uncertainty about economic and general environment. With economic growth, uncertainty is also increasing. This makes each decision costlier and time-consuming. OR is, thus, quite essential from reliability point of view.

(d) *knowledge explosion*: knowledge is increasing at a very fast rate. Majority if the industries are not up-to-date with the latest knowledge and are, therefore, at a disadvantage. OR teams collect the latest information for analysis purposes which is quite useful for the industries.

1.5 SCOPE OF OPERATIONS RESEARCH

Having known the definition of OR, it is easy to visualize the scope of operations research. When we broaden the scope of OR, we find that it has really been practised for hundreds of years even before world war II. Whenever there is a problem of optimization, there is the scope for the application of OR. Its techniques have been used in a wide range of situations:

1. In Industry

In the field of *industrial management*, there is a chain of problems starting from the purchase of raw materials to the dispatch of finished goods. The management is interested in having an overall view of the method of optimizing profits. In order to take a decision on scientific basis, OR team will have to consider various alternative methods of producing the goods and the return in each case. OR study should also point out the possible changes in the overall structure like installation of new machine, introduction of more automation, etc. OR has been successfully applied in industry in the fields of production, blending, product mix, inventory control, demand forecast, sale and purchase, transportation, repair and maintenance, scheduling and control of projects and scores of other associated areas.

2. In Defence

OR has a wide scope for application in *defence operations*. In modern warfare the defence operations are carried out by a number of different agencies, namely airforce, army and navy. The activities performed by each of them can be further divided into sub-activities viz. operations, intelligence, administration, training and the like. There is thus a need to coordinate the various activities involved in order to arrive at optimum strategy and to achieve consistent goals. Operations research, conducted by team of experts from all the associated fields, can be quite helpful to achieve the desired results.

3. Planning

In both *developing and developed economics*, OR approach is equally applicable. In developing economies, there is a great scope of developing an OR approach towards *planning.*

The basic problem is to orient the planning so that there is maximum growth of per capita income in the shortest possible time, by taking into consideration the national goals and restrictions imposed by the country. The basic problem in most of the countries inasia and Africa is to remove poverty and hunger as quickly as possible. There is, therefore, a great scope for economist, statisticians, administrators, technicians, politicians and agriculture experts working together to solve this problem with an OR approach.

4. Agriculture

OR approach needs to be equally developed in *agriculture sector* on national or international basis. With population explosion and consequent shortage of food, every country is facing the problem of optimum allocation of land to various crops in accordance with climatic conditions and available facilities. The problem of *optimal distribution of water from thevarious water resources* is faced by each developing country and a good amount of scientific work can be done in this direction.

5. Public Utilities

OR methods can also be applied in big *hospitals* to reduce waiting time of out-door patients and to solve the administrative problems.

Monte Carlo methods can be applied in the area of *transport* to regulate train arrivals and their running times. Queuing theory can be applied to minimize congestion and passengers waiting time.

OR is directly applicable to *business and society*. For instance, it is increasingly being applied in *L.I.C. offices* to decide the premium rates of various policies. It has also been extensively used in petroleum, paper, chemical, metal processing, aircraft, rubber, transport and distribution, mining and textile industry.

OR approach is equally applicable to big and small organizations. For example, whenever a departmental store faces a problem like employing additional sales girls, purchasing an additional van, etc., techniques of OR can be applied to minimize cost and maximize benefit for each such decision.

Thus we find that OR has a diversified and wide scope in the social, economic and industrial problems of today.

1.6 OPERATIONS RESEARCH AND DECISION-MAKING

Operations research or management science, as the name suggests, is the science of managing. As is known, management is most of the time making decisions.

Operations research provides the data to analyst to take the decision and means to apply scientific, systematic, technical and mathematical methods for taking the appropriate decision and solve the problem. OR does not provide decisions it only present quantitative data to the managers. The managers use this quantitative data for taking the decisions and find out the better decision. Hence, it is used to solve complex problems. OR helps to take decisions about operation and production Operations research (OR) is an analytical, logical and systematic method of problem-solving and decision-making that is helpful in the management of organizations. In operations research, problems are broken down into basic components and then solved in defined steps by mathematical analysis.Analytical methods used in OR include mathematical logic, simulation, network analysis, queuing theory and game theory etc. Meaning and definition It is the combination of two terms

Operational + Research Operational:- working as in order, ready for use Research:- the systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions.

Operational Research:-the application of scientific and mathematical techniques to the study and analysis of problems is called operational research which provide us the data to take the decision. Achievement by Using of Operational Research

Forecasting and Scheduling:- OR Methods and models provide data to airlines vehicles in supply chains, orders in a factory and of operating theatres in a hospital to take the decision.

Capacity planning and facility planning:- computer simulations of airports for the rapid and safe processing of travelers, improving appointments systems for medical practice. Planning and developing: OR recognize the possible future developments in telecommunications and help in deciding how much capacity and competence is needed in a business.

Capitulate management:- OR helps in setting the cost of airlines space and seat and hotel rooms. So that changes can take place as per demand and probable risk can be taken by the producer Credit scoring: OR Helps to the finance company to decided about customers and offer the best prospects for credit companies.

Sales promotion:- OR helps in evaluate and estimate the value of sale promotions, marketing finding developing customer profiles and computing the life-time value of a customer.

Advantages of Operations Research in Decision Making Efficient Control: The management of organizations recognizes that it is a complicated and expensive issue that why it needs proper supervision and control and to execute in the organization it needs continuous executive supervision to every routine work. An O.R. approach may provide the manager an analytical, logical andquantitative basis to identify the area of problem. The major area in this category is production, manufacturing and construction scheduling and inventory replacement.

Effective Systems: Operational research techniques are to analyze the problem of decision making such as good site for plant , whether to open a new storehouse, etc. It also helps in assortment of cost-effective means of transportation, jobs sequencing, production scheduling, replacement of old machinery, etc. Better and Analytical Decisions Operational Research Techniques and models helpful to make the better decisions and diminish the risk of making flawed decisions. These methodology gives the executive an better foresight how to makes the decisions. Good Co-ordination and Management An operations-research-oriented techniques are useful for an effective and efficient planning model that helps in co-coordinating relationship between the different divisions of a company . Maximize profits & Minimize Losses The operational Research techniques provide the quantitative data to take the decision which help the analyzer to take the appropriate decision which lead to maximize the profit and minimize the losses. Limitations of Operations Research in decision making Operated by Technical & Electronic Devices: These days operational research techniques obtain an optimal solution using various computational systems, model and techniques. In this scenario , these factors are colossal and expressing them in quantity and establish the relationships among these require calculations that can only be handled by computers.

Solve only quantitative problems: O.R. techniques provide a solution only when all the elements related to a problem can be quantified. All relevant variables do not lend themselves to quantification. Factors that cannot be quantified find no place in O.R. models.

Widen the Gap between executive and Researcher: O.R. being specialist's job requires a mathematician or a statistician, who might not be aware of the business problems. Similarly, a manager fails to understand the complex working of O.R. Thus, there is a gap between the two. More time and cost consuming: When the basic data are subjected to frequent changes, incorporating them into the O.R. models is a costly affair. Moreover, a fairly good solution at present may be more desirable than a perfect O.R. solution available after sometime. Problem in execution and Implementation: Implementation of decisions is a delicate task. It must take into account the complexities of human relations and behavior Role of Operations Research in decision Making Operations research plays an important role in almost all areas of decision making.

Some areas where operational research (OR) techniques can be used as listed below:

Financing and Investment Polices: This category includes the analysis of credit and loan policy, fund flow and cash flow and also considers the dividend, share and bonus policy. It reflects the portfolio of investment too. Selling, Promotion, Marketing & Publicity It emphasis on selection of the product and timing which means what type of commodity should be manufactured at the proper passage of the time. It focuses on the media of publicity like print media and electronic media may be chosen. It also used to know about how many number of sales persons are required to some specific area. More over OR also use to select the product mixed which include product, price, place and promotions. Acquisition, Investigation, Manufacturing and Personnel Management The operational research techniques being used to purchase, substitute and reorganize the most advantageous policies. These techniques also include location and manufacturing size of retail outlets, factories and warehouses, loading and unloading facilities for trucks, allocation and scheduling of resources and optimum to merge the products, recruitment and selection of best suitable employee for a company and job assignments etc.Research and Development The methods currently available in literature for numerically solving operational research problems may be broadly classified as deterministic methods and probabilistic methods. The deterministic methods try to guarantee that a neighborhood of the solution of a problem is attained. Such methods do not use any stochastic techniques, but rely on a thorough search of the feasible domain. They are applicable, however, to a restricted class of problems only. On the other hand the probabilistic methods make use of probabilistic or stochastic approach to search for the solutions for the problems. Although probabilistic methods do not give an absolute guarantee of the exact solution having been obtained, these methods are sometimes preferred over the deterministic methods because they are applicable to wider class of problems. There does not exist any operational research technique which can say I am the best. Therefore there is always a need to investigate and develop an operational research technique for better solution. Scope of Operation Research in Decision Making OR has used effectively in many area of research. The Operation Research may be considered as a tool which is employed to raise the efficiency of management decisions. Scientific method of OR is used to comprehend and explain the phenomena of operating system.It is useful in the following various important fields In agriculture and cultivation As we seen that the population is increasing day by day and resultant of it there is deficiency of groceries and every country of the world facing the problem of most favorable or best possible allocation of the land for their crops as it is requires by the climate conditions. Moreover the problem of water resources distribution and allotment for the same is also the problem of developing countries. So the operational research techniques provide the data to determine the policy and by effects of the policy proper action can take place in the better directions. In Economy It is vital need for every government to do a careful planning for the economic progress of the country. So that the country should be prepare to face the economic crisis by using the operational research methodology an institute can establish the profit plan. It is also apply to increase the per capital income by using the limited recourses. The replacement and alternate polices is depend on the OR techniques In corporate house and business As we know that experience makes the man perfect and the executives of the industry take the decision by using its past experience but at the age of retirement or due any reason he left the job the loss can be stumble upon by the organization. Thus apply the Operational techniques to the business an executive can take the right and appropriate decision about using the limited resources of land, labor, machinery, capital and other resources of the production In promotion and selling of the product OR techniques assist the marketing executives to take the decision about where they should sell their product, how they can transport the product at minimum cost. It also helps in decide about the price of the commodity, requirement of the stock as per the future demand of the product. By using the appropriate OR technique there can be proper choice of advertise media as per cost constraint which type of media would be economical print or electronic media.

In human resources management A personnel manager can make use of OR techniques to appoint the highly suitable and qualified human being on minimum salary, OR technique help to calculate the retirement age of the staff. Moreover OR Techniques help t in recruitment of person as per the requirement of the organization whether on contract basis or permanent basis, fulltime or seasonal working In production and manufacturing OR technique helps the production manager to calculate the size and how many items should be produced. It is helpful in scheduling and sequencing of the machine. OR methodology helps to choose the appropriate location of the plant, plant design etc. In Life Insurance Plan and Premium. OR approach is also applicable to facilitate the L.I.C offices to decide What should be the premium rates for a range of policies.

Phases in Operation Research Study:

Since, the main objective of operation research is to provide better quantitative information's for making decision. Now our aim is to learn how we can have better decisions.

The procedure for making decisions with the OR study generally involves the following phases:

(i) Judgment Phase:

i. Determination of operation.

ii. Determination of objectives.

iii. Determination of effectiveness of measures.

iv. Determination of type of problem, its origin and causes.

(ii) Research Phase:

i. Observation and data collection for better understanding of the problem.

ii. Formulation of relevant hypothesis and models.

iii. Analysis of available information and verification of hypothesis.

iv. Production and generation of results and consideration of alternatives.

(iii) Action Phase:

i. Recommendations for remedial action to those who first posed the problem, this includes the assumptions made, scope and limitations, alternative courses of action and their effect.

ii. Putting the solution to work: implementation.

Without OR, in many cases, we follow these phases in full, but in other cases, we leave important steps out. Judgment and subjective decision-making are not good enough. Thus industries look to operation research for more objective way to make decisions. It is found that method used should consider the emotional and subjective factors also.

For example, the skill and creative labour are important factors in our business and if management wants to have a new location, the management has to consider the personal feeling of the employees for the location which he chooses.

Scope of Operation Research:

In its recent years of organised development, O.R. has solved successfully many cases of research for military, the government and industry. The basic problem in most of the developing countries in Asia and Africa is to remove poverty and hunger as quickly as possible. So there is a great scope for economist, statisticians, administrators, politicians and technicians working in a team to solve this problem by an O.R. approach.

On the other hand, with the explosion of population and consequent shortage of food, every country is facing the problem of optimum allocation of land for various crops in accordance with climatic conditions and available facilities. The problem of optimal distribution of water from a resource like a canal for irrigation purposes is faced by developing country. Hence a good amount of scientific work can be done in this direction.

In the field of Industrial Engineering, there is a claim of problems, starting from the procurement of material to the despatch of finished products. Management is always interested in optimizing profits.

Hence in order to provide decision on scientific basis, O.R. study team considers various alternative methods and their effects on existing system. The O.R. approach is equally useful for the economists, administrators, planners, irrigation or agricultural experts and statisticians etc.

Operation research approach helps in operation management. Operation management can be defined as the management of systems for providing goods or services, and is concerned with the design andoperation ofsystems for the manufacture, transport, supply or service. The operating systems convert the inputs to the satisfaction of customers need.

Thus the operation management is concerned with the optimum utilisation of resources i.e. effective utilisation of resources with minimum loss, under utilisation or waste. In other words, it is concerned with the satisfactory customer service and optimum resource utilisation. Inputs for an operating system may be material, machine and human resource.

O.R. study is complete only when we also consider human factors to the alternatives made available. Operation Research is done by a team of scientists or experts from different related disciplines.

For example, for solving a problem related to the inventory management, O.R. team must include an engineer who knows about stores and material management, a cost accountant a mathematician-cum-statistician. For large and complicated problems, the team must include a mathematician, a statistician, one or two engineers, an economist, computer programmer, psychologist etc.

Some of the problems which can be analysed by operations research are given hereunder:

1. Finance, Budgeting and Investment:

i. Cash flow analysis, long range capital requirement, investment portfolios, dividend policies,

ii. Claim procedure, and

iii. Credit policies.

2. Marketing:

i. Product selection, competitive actions,

ii. Number of salesmen, frequencies of calling on, and

iii. Advertising strategies with respect to cost and time.

3. Purchasing:

i. Buying policies, varying prices,

ii. Determination of quantities and timing of purchases,

iii. Bidding policies,

iv. Replacement policies, and

v. Exploitation of new material resources.

4. Production Management:

i. Physical distribution: Location and size of warehouses, distribution centres and retail outlets, distribution policies.

ii. Facilities Planning: Number and location of factories, warehouses etc. Loading and unloading facilities.

iii. Manufacturing: Production scheduling and sequencing stabilisation of production, employment, layoffs, and optimum product mix.

iv. Maintenance policies, crew size.

v. Project scheduling and allocation of resources.

5. Personnel Management:

i. Mixes of age and skills,

ii. Recruiting policies, and

iii. Job assignments.

6. Research and Development:

i. Areas of concentration for R&D.

ii. Reliability and alternate decisions.

iii. Determination of time-cost trade off and control of development projects.

Simplex method in operations research:

Simplex method is an iterative procedure that allows to improve the solution at eachstep. This procedure is finished when isn't possible to improve the solution.

Starting from a random vertex value of the objective function, Simplex method tries to find repeatedly another vertex value that improves the one you have before. The search is done through the side of thepolygon (or the edges of the polyhedron, if the number of variables is higher). As the number of vertices (and edges) is finite, it will always be able to find the result.

Simplex method is based on the following property: if objective function, F, doesn't take the max value in the A vertex, then there is an edge starting at A, along which the value of the function grows.

You should take care about Simplex method only works with "≤" type inequality and independent coefficients higher or equal to zero, and you will have to standardize the restrictions for the algorithm. Case after this procedure "≥" o "=" type restrictions appear (or not modified) you should try other ways, being Two-Phase Simplex method the best choice.

Preparing the model to adapt it to the Simplex method

This is the standard way of the model:

Objective function:	c1·x1 + c2·x2 + ... + cn·xn
Subject to:	a11·x1 + a12·x2 + ... + a1n·xn = b1
	a21·x1 + a22·x2 + ... + a2n·xn = b2
	...
	am1·x1 + am2·x2 + ... + amn·xn = bm
	x1,..., xn ≥ 0

To do this you must follow these rules:

1. The objective must be maximize or minimize the function.
2. All restrictions must be equal.
3. All variables are not negatives.
4. The independent terms are not negatives.

Changing the optimization type.

If we want to minimize our model, we can keep it, but we must consider the new criteria for the halt condition (stop iterations when all coefficients in the value objective function row are less or equal to zero), and the leaving condition row. In order to not change criteria, we can convert the minimize objective function F to maximize objective F·(-1).

Advantages: We will not have to worry about halting criteria, or exit condition of rows, since they keep on.

Inconveniences: In the event of the function have all his basic variables positive, and further the restrictions are inequality "≤", they become negative when doing the change and plus signs remain in the row of the value of the objective function, then Simplex method obeys the halting condition, and that optimal value would be obtained is 0, by default.

Solution: In fact this kind of problem does not exist, since so that the solution is greater than 0, any restriction should have the condition "≥", and then we would go into a model for the Two-Phase Simplex method.

Converting the independent term sign (constants to the right of restrictions)

We will have to arrange our model so that the independent terms of restrictions will be greater or equal to 0, if not, Simplex method cannot be used. The only thing that would be necessary to do is multiply by "-1" the restrictions where independent terms be less than 0.

Advantages: With this simple modification of signs in restriction we can use Simplex method.

Inconveniences: It can work out in restrictions where we have to modify the signs of constants, the signs of inequalities be ("=", "≤"), becoming ("=","≥") what in any event we will have to develop the Two-Phase Simplex method. This inconvenience is not controllable, although it would be able to benefit us only if terms of inequality exist ("≤","≥"), and terms "≥" coincide with restrictions where the independent term is negative.

Normalization restrictions:

If an inequality of the type "≥", appears in our model, will we have to add a new variable, called surplus variable *si*, with restriction si ≥ 0. The new variable appears with coefficient equal to zero in the objective function, and subtracting in inequalities.

A problem appears to us, let's see how to solve inequalities that contains an inequality type "≥" :

$$a11 \cdot x1 + a12 \cdot x2 \geq b1 \longrightarrow a11 \cdot x1 + a12 \cdot x2 - 1 \cdot xs = b1$$

As all our models based on that all his variables are greater or equal than zero, when we do the first iteration in the Simplex's model, the basic variables will not be in the base and they will take value zero, and all others will maintain their values. In this case our variable xs, after doing zero to x1 and x2, will take the value - b1. The condition of not negativeness will not come true, so it will be necessary to add a new variable, xr, that will appear in the objective function with zero coefficient, and adding in the inequality of correspondent restriction. Would be left of the following way:

$$a11 \cdot x1 + a12 \cdot x2 \geq b1 \longrightarrow a11 \cdot x1 + a12 \cdot x2 - 1 \cdot xs + 1 \cdot xr = b1$$

This type of variables are called artificial variables, and they will appear when there are inequalities with inequality ("=","≥"). This will take us compulsorily to accomplish the Two-Phase Simplex method, that will explain later on.

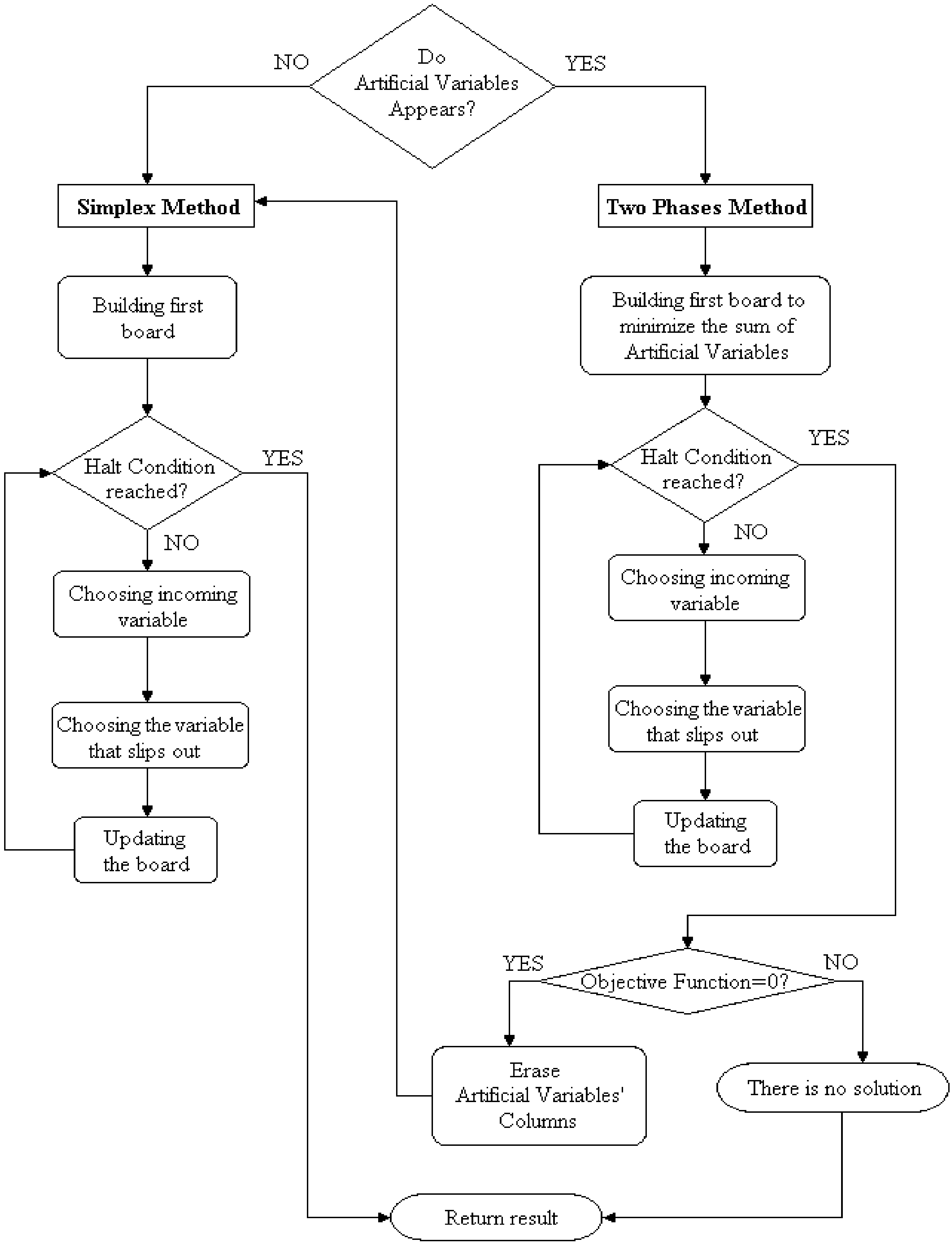
Itself mode, if inequality has "≤" type, we will have to add a new variable, called slack variable si, with restriction si "≥" 0. The new variable appears with zero coefficient in the objective function, and adding up in the inequalities.

To sum up we can let this board, according to the inequality that appears, and with the value that the new variables must be with.

Type of inequality	Type of variable
≥	- surplus + artificial
=	+ artificial
≤	+ slack

Starting the Simplex method

Once we have standardized our model, it can happen to go into the Simplex method or Two-Phase Simplex method. See yourself in the figure as we must perform on for reach the solution of our problem.


Do
Artificial Variables
Appears?
NO
YES
Simplex Method
Two Phases Method
Building first
board
Building first board to
minimize the sum of
Artificial Variables
Halt Condition
reached?
YES
NO
Choosing incoming
variable
Choosing the variable
that slips out
Updating
the board
Halt Condition
reached?
YES
NO
Choosing incoming
variable
Choosing the variable
that slips out
Updating
the board
Objective Function=0?
YES
NO
Erase
Artificial Variables'
Columns
There is no solution
Return result

Simplex method- First Board Construction:

In the board's first column will appear that we will call base, in the second one, the coefficient that each variable that appears at base has in the objective function (we will call this column Cb), in third column, the independent term of every restriction (P0), and from this column will appear each variable of the objective function (Pi). In order to have a more obvious vision of the board, we will include a row that we will put each one of the names of columns in. On this board we have, we will include two new rows: One that will lead the board, where the constants of the coefficients of the objective function will appear, and another one that will be the last row, where the objective function will take value. Our final board will have such rows as restrictions.

Tableau						
			C_1	C_2	...	C_n
Base	**C_b**	**P_0**	**P_1**	**P_2**	**...**	**P_n**
P_1	C_{b1}	b_1	a_{11}	a_{12}	...	a_{1n}
P_2	C_{b2}	b_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...	...	...
P_m	C_{bm}	b_m	a_{m1}	a_{m2}	...	a_{mn}
Z		Z_0	Z_1-C_1	Z_2-C_2	...	Z_n-C_n

Z row's values are obtained this way: Z0 value will be the result of substituting Cim in the objective function (zero else appears in the base). The left columns are obtained subtracting to this value the one belonging to the coefficient that appears in the board's front row.

It will be observed when realizing Simplex method, that slack variables will be in the base, in this first table.

- Halt Condition: Will check if we must do a new iteration or not to do, that it will be know if in Z row appears any negative value. If this is not the case, it means we have reached the problem's optimal solution.

- Choosing incoming variable: If halt condition has not come true, we must choose one variable to enter the base in the next board. For it we look for strictly negative values of the Z row, and the minor will be which give us the incoming variable.

- Choosing the variable that slips out: Once we have obtained the incoming variable, the coming out variable will be reached, having nothing else to do that select that row whose quotient P0/Pj be the lowest among strictly positives (considering that only will be done when Pj be greater than 0). The intersection among incoming column and the coming out row will determine us the pivot element.

- Updating the board: The correspondent rows to the objective function and titles will remain unaltered in the new board. Left rows will be calculated with two ways:

- If we are trying with pivot row, each element will result from:

New Pivot Row Element = Actually Pivot Row Element / Pivot.

- The left elements of rows will be reached so:

*New Row Element = Actually Pivot Row Element - (Pivot Column Element from actually Row * New Row Element).*

Two-Phase Simplex method

This method differs from Simplex method that first it is necessary to accomplish an auxiliary problem that has to minimize the sum of artificial variables. Once this first problem is resolved and reorganizing the final board, we start with the second phase, that consists in making a normal Simplex.

1st Phase

At this first phase, all can be done like in Simplex method, except the first board's construction, halt condition and preparing the board that will be used in the second phase.

- First Board Construction: We proceed in same way as Simplex method, but with some differences. Objective Function row is different in the first phase, because the objective function changes, due to it will appear every term with zero value, but them which are artificial variables, which have "-1" value because we are minimizing the sum of this variables (remember that minimize F is the same that maximize F·(-1)).

The other difference for this first board consists in the way of calculating the row Z. It will have to be calculated the following way: The Cb·Pjproducts will be added for all rows and to this sum we must subtract the value that appears (according to the column that we are doing) in the objective function row.

Tableau								
		C_0	C_1	C_2	...	C_{n-k}	...	C_n
Base	**C_b**	**P_0**	**P_1**	**P_2**	**...**	**P_{n-k}**	**...**	**P_n**
P_1	C_{b1}	b_1	a_{11}	a_{12}	...	a_{1n-k}	...	a_{1n}
P_2	C_{b2}	b_2	a_{21}	a_{22}	...	a_{2n-k}	...	a_{2n}
...	...	...	...	...	...	...	...	...
P_m	C_{bm}	b_m	a_{m1}	a_{m2}	...	a_{mn-k}	...	a_{mn}
Z		Z_0	Z_1	Z_2	...	Z_{n-k}	...	Z_n

Being $Zj = \Sigma(Cb \cdot Pj) - Cj$ with $Cj = 0$ for all decision, slacks and surplus variables and $Cj = -1$ for artificial variables.

- Halt condition: The halt condition is the same that in Simplex method. The difference reside in that can occur two cases when halt condition is reached: the function takes zero value, it means that the original problem has solution, or function takes a different value, suggesting that our model does not have solution.

- Erasing artificial variables columns: If we have reached the conclusion that the original problem has solution, we must prepare our board for the second phase. The artificial variables columns will be erased, modify the objective function row instead original, and calculate Z row at same way that in the 1st phase's first board.

Noticing anomalous cases and solutions

Obtaining the solution: When the halt condition is reached, you can see the values of the basic variables which are in the base and the optimal value that the function takes, looking at P0 column. At case you're minimizing, this optimal value must be multiplied by "-1".

Infinite solutions: Once the halt condition is obeyed, if you notice that any variable that doesn't appear in the base, has a 0 value at row Z, it means there are other solutions that give you the same optimal value for the objective function. This is a problem which admits infinite solutions, all them among the segment (or plane portion, or space region, etc. depending on the number of variables) that defines Ax+By=Z0. You could do more iterations using as incoming variable any of the variables in the Z row which have zero value, and you would have other solutions.

Unbounded solution: When you are searching the outgoing variable whether you notice that every variable in the incoming variable column have all their elements negative or void, it's a problem which has an unbounded solution. So there is no optimal concrete value. If the values of the variables grow, the objective function value also grows without violates any restriction.

Solution does not exist: In case that there seems to be no solution, we will have to solve it using Two-Phase Simplex method, so at the end of the 1st phase we will know if we are in such situation.

Tie of incoming variable: You can choose anyone of them, unless it affect the final solution, the inconvenience that it presents is that according to this choose you will have to do more or less iterations. Is counseled to choose to favor of the basic variables, since are those that will stay on the base at the end of the method.

Tie of coming out variable: Again you can choose anyone of them, although it can occur degenerated case and entering into loop cycles. In order to avoid them as far as possible, we will have prejudice in favor of basic variables doing that they remain in the base. At the case to be in the first phase (of the Two-Phase Simplex method), we will choose in case of tie to take out the artificial variables.

Curiosity in the 1st Phase: When the first phase finalizes, if the original problem has solution all the artificial variables in the Z row must have value "1".

Can the pivot be 0?: It can not be 0, because, quotients must be greater than 0.

LINEAR PROGRAMMING

A linear programming problem may be defined as the problem of maximizing or minimizing a linear function subject to linear constraints. The constraints may be equalities or inequalities. Here is a simple example. Find numbers x1 and x2 that maximize the sum $x1 + x2$ subject to the constraints $x1 \geq 0$, $x2 \geq 0$, and $x1 + 2x2 \leq 4$ $4x1 + 2x2 \leq 12$ $-x1 + x2 \leq 1$ In this problem there are two unknowns, and five constraints. All the constraints are inequalities and they are all linear in the sense that each involves an inequality in some linear function of the variables. The first two constraints, $x1 \geq 0$ and $x2 \geq 0$, are special. These are called non-negativity constraints and are often found in linear programming problems. The other constraints are then called the main constraints. The function to be maximized (or minimized) is called the objective function. Here, the objective function is $x1 + x2$. Since there are only two variables, we can solve this problem by graphing the set of points in the plane that satisfies all the constraints (called the constraint set) and then finding which point of this set maximizes the value of the objective function. Each inequality constraint is satisfied by a half-plane of points, and the constraint set is the intersection of all the half-planes. In the present example, the constraint set is the fivesided figure shaded in Figure 1. We seek the point (x1, x2), that achieves the maximum of $x1 + x2$ as (x1, x2) ranges over this constraint set. The function x1 + x2 is constant on lines with slope −1, for example the line x1 + x2 = 1, and as we move this line further from the origin up and to the right, the value of x1 + x2 increases. Therefore, we seek the line of slope −1 that is farthest from the origin and still touches the constraint set. This occurs at the intersection of the lines x1 + 2x2 = 4 and 4x1 + 2x2 = 12, namely, (x1, x2) = (8/3, 2/3). The value of the objective function there is (8/3) + (2/3) = 10/3. Exercises 1 and 2 can be solved as above by graphing the feasible set. It is easy to see in general that the objective function, being linear, always taken its maximum (or minimum) value at a corner point of the constraint set.

Standard form

Standard form is the usual and most intuitive form of describing a linear programming problem. It consists of the following three parts:

(i) A linear function to be maximized
(ii) Problem constraints of the following form
(iii) Non variable

The problem is usually expressed in *matrix form*, and then becomes:
Other forms, such as minimization problems, problems with constraints on alternative forms, as well as problems involving negative variables can always be rewritten into an equivalent problem in standard form.

DUAL SIMPLEX METHOD

The **Dual Simplex Method** offers an alternative when solving Linear Programming (LP) models with algorithms. This method may be used in particular when the standard way to carry a linear programming model is not available from an initial **basic feasible solution**. Consider the following LP problem to illustrate the application of the Dual Simplex Method:

$$Min \quad 160X_1 + 120X_2 + 280X_3$$

$$\text{s.a.} \quad 2X_1 + X_2 + 4X_3 \geq 1$$

$$2X_1 + 2X_2 + 2X_3 \geq 3/2$$

$$X_1 \geq 0, X_2 \geq 0, X_3 \geq 0$$

To solve the problem above in a standard form, we must add 2 excess non-negative variables for constraints 1 and 2, which we will respectively call X4 and X5.

$$Min \quad 160X_1 + 120X_2 + 280X_3$$

$$\text{s.a.} \quad 2X_1 + X_2 + 4X_3 - X_4 = 1$$

$$2X_1 + 2X_2 + 2X_3 - X_5 = 3/2$$

$$X_i \geq 0, i = 1,2,3,4,5$$

Then we build the initial tableau of the Simplex Method:

X1	X2	X3	X4	X5	
2	1	4	-1	0	**1**
2	2	2	0	-1	**3/2**
160	**120**	**280**	**0**	**0**	**0**

How to continue the iterations of the Simplex Method?

Before that, it is necessary to have an initial basic feasible solution. In this context if we want to use X4 and X5 as basic variables (and hence X1, X2 and X3 as basic variables) it is required that X4 and X5 are greater than or equal to zero, however, their coefficients in the respective rows are negative and therefore we can't use them (matrix with "1" as a diagonal, and all the other coefficients at zero). So to form the identity we can multiply by ***"-1"*** row 1 and 2, obtaining the following:

X1	X2	X3	X4	X5	
-2	-1	-4	1	0	**-1**
-2	-2	-2	0	1	**-3/2**
160	**120**	**280**	**0**	**0**	**0**

Now X4 and X5 are basic variables and adopt the values of -1 and -3/2, respectively, which clearly does not satisfy the conditions of non-negativity for the decision variables, i.e. they do not correspond to a basic feasible solution. However, in this instance we can apply the **Dual Simplex Method** as an alternative resolution. To do this, we will select a variable to leave the base and adopt as a criterion for basic variable associated with **"more negative"** **right hand side** **(RHS)**. In this instance the variable is X5. Then to determine which variable will go into the base we find a minimum quotientbetween the negative of the reduced cost of the non-basic variables and the strictly less than zero entries for non-core variables in row 2 (row associated with more negative right). I.e.: **Min {-160/-2; -120/-2; -280/-2} = 60**, then the minimum quotient is reached in the second column associated with the non-basic variable X2, therefore said variable enters the base.

SENSITIVITY ANALYSIS:

Sensitivity Analysis deals with finding out the amount by which we can change the input data for the output of our linear programming model to remain comparatively unchanged. This helps us in determining the sensitivity of the data we supply for the problem. If a small change in the input (for example in the change in the availability of some raw material) produces a large change in the optimal solution for some model, and a corresponding small change in the input for some other model doesn't affect its optimal solution as much, we can conclude that the second problem is more robust then the first. The second model is less sensitive to the changes in the input data.

We will consider the case of sensitivity of the optimum solution to changes in the availability of the resources. (Right hand side of the constraints.)

If in any linear programming problem there are n variables and m constraints we can think as the right hand sides as being the representatives of the amount of resources. For example, consider our old chemical company model:

Maximize z =

subject to,

The right hand sides represented various resources: the amount of raw materials 1 and 2, the market limit, and the daily demand. Now if these right hand sides were changed the whole problem would change. Suppose we want to know is the worth of any particular resource. More precisely, we want to know how much is the amount of the first raw material being 24 units available really important. If we increase the amount from 24 to 25 our optimal value changes to 21.75, whereas previously it was 21. So a unit increase in the amount of the first resource changes the optimal value (which is the total profit) by 0.75. This can be therefore thought of as the *unit worth* of the first resource. The technical term for this is the first

resources' *shadow price*

Determination of shadow prices

To determine the shadow prices for the resources individually by increasing them one by one and solving the associated linear models is extremely inefficent. Let us examine another way of calculating the shadow prices.

Consider the following linear system:

Maximize z =

subject to,

The optimal simplex table with slack variables is:

Basic							BFS
Z	4	0	0	1	2	0	1350
		1	0			0	100
		0	1	0		0	230
	2	0	0	-2	1	1	20

If resource 1 is increased by one unit it means that the first slack variable is decreased by one unit so that the equality between the left and the right hand sides remains unchanged. A similar thing holds for all other resources as well.

Goal programming

It is a branch of multiobjective optimization, which in turn is a branch of multi-criteria decision analysis (MCDA). This is an optimization programme. It can be thought of as an extension or generalisation of linear programming to handle multiple, normally conflicting objective measures. Each of these measures is given a goal or target value to be achieved. Unwanted deviations from this set of target values are then minimised in an achievement function. This can be a vector or a weighted sum dependent on the goal programming variant used. As satisfaction of the target is deemed to satisfy the decision maker(s), an underlying satisficing philosophy is assumed. Goal programming is used to perform three types of analysis:

1. Determine the required resources to achieve a desired set of objectives.
2. Determine the degree of attainment of the goals with the available resources.
3. Providing the best satisfying solution under a varying amount of resources and priorities of the goals

VARAINTS

The initial goal programming formulations ordered the unwanted deviations into a number of priority levels, with the minimisation of a deviation in a higher priority level being infinitely more important than any deviations in lower priority levels. This is known as *lexicographic* or pre-emptive goal programming. Ignizio[4] gives an algorithm showing how a lexicographic goal programme can be solved as a series of linear programmes. Lexicographic goal programming should be used when there exists a clear priority ordering amongst the goals to be achieved.

If the decision maker is more interested in direct comparisons of the objectives then *Weighted* or non pre-emptive goal programming should be used. In this case all the unwanted deviations are multiplied by weights, reflecting their relative importance, and added together as a single sum to form the achievement function. It is important to recognise that deviations measured in different units cannot be summed directly due to the phenomenon of incommensurability.

Hence each unwanted deviation is multiplied by a normalisation constant to allow direct comparison. Popular choices for normalisation constants are the goal target value of the corresponding objective (hence turning all deviations into percentages) or the range of the corresponding objective (between the best and the worst possible values, hence mapping all deviations onto a zero-one range).[6] For decision makers more interested in obtaining a balance between the competing objectives, *Chebyshev* goal programming should be used.

Introduced by Flavell in 1976,[10] this variant seeks to minimise the maximum unwanted deviation, rather than the sum of deviations. This utilises the Chebyshevdistancemetric, which emphasizes justice and balance rather than ruthless optimisation.

STRENGTHS AND WEAKNESS:

A major strength of goal programming is its simplicity and ease of use. This accounts for the large number of goal programming applications in many and diverse fields. Linear Goal programmes can be solved using linear programming software as either a single linear programme, or in the case of the lexicographic variant, a series of connected linear programmes.

Goal programming can hence handle relatively large numbers of variables, constraints and objectives. A debated weakness is the ability of goal programming to produce solutions that are not Pareto efficient. This violates a fundamental concept of decision theory, that is no rational decision maker will knowingly choose a solution that is not Pareto efficient. However, techniques are available [6][11][12] to detect when this occurs and project the solution onto the Pareto efficient solution in an appropriate manner.

The setting of appropriate weights in the goal programming model is another area that has caused debate, with some authors [13] suggesting the use of the Analytic Hierarchy Process or

interactive methods [14] for this purpose

1.11 DIFFICULTIES IN OPERATIONS RESEARCH

The previous sections have brought out the positive side of OR only. However, there is also the need to point out the negative side. Certain common traps and pitfalls can have, ruined the otherwise good work. Some of these pitfalls are quite obvious while others are so subtle and hidden the extreme care is required to locate their presence.

In the very first phase of OR—the problem formulation phase—a number of pitfalls can and do arise. It is necessary that the right problem be selected and it must be completely and accurately defined. Is the right problem being solved?Is the scope considered wide and proper? Will it result in optimization or only sub-optimization? Will the solution properly reflect the objectives as well as the imposed constraints? Are proper effectiveness measures being used? This phase of problem formulation is perhaps the most important and toughest part of OR study.

Secondly, data collection may also consume a very large portion of time and money spent on OR study.

Thirdly, the whole study by operations analyst is based on his observations in the past. Strictly speaking, these observations can only be related to the laws that operated in the past, as there is no evidence that the laws will continue to operate in future also. If the laws are applied to the future, it clearly amounts to extrapolation in time.

Fourthly, the operations researchers, while making observations, may affect the behaviour of the system he is studying. Moreover, however comprehensive his experiments may be, his observation can never be more than a sample of the whole. These difficulties present special hazards to operations researchers. His aim is to find out what happens in a working organization. He can get the information in two ways: by direct observations or from the previous records. The behavior of an organization depends upon the activities of the persons in it and the very fact that they are being observed is bound to affect their behavior. On the other hand, accuracy of previous records is always doubtful and they seldom provide the complete information in all the points sought.

Perhaps the greatest difficulty in OR, however, is created by the time factor. The managers have to make decisions one way or the other, and a fairly good solution to the problem at the right time may be much more useful than the perfect solution too late. Further, the cost involved is also an important factor. Sometimes, some simple application of OR may yield a good solution quickly and it may be unwise to spend a lot of money and effort to produce a slightly better solution much later.

Other pitfalls in problem solving include:

(i) warping the problem to fit a standard model, tool or technique,

(ii) failure to test the model and solution before implementation, and

(iii) failure to establish proper controls.

Lastly, what may appear to be a pitfall is the fact that OR study may raise more questions than it answers. However, this may ultimately result in more deep insight into the system, yielding further benefits and improvements.

1.12 LIMITATIONS OF OPERATIONS RESEARCH

1. Mathematical models, which are essence of OR, do not take into account qualitative factors or emotional factors which are quite real. All influencing factors which cannot be quantified find no place in mathematical models.

2. Mathematical models are applicable to only specific categories of problems.

3. OR tries to find optimal solution taking all the factors of the problem into account. Present problems involve numerous such factors; expressing them in quantity and establishing relations among them requires huge calculations.

4. Being a new field, generally there is a resistance from the employees to the new proposals.

5. Management, who has to implement the advised proposals, may itself offer a lot of resistance due to conventional thinking.

6. Young enthusiasts, overtaken by its advantages and exactness, generally forget that OR is meant for men and not that men are meant for it.

Thus at the implementation stage, the decision cannot be governed by quantitative considerations alone. It must take into account the delicacies of human relationships. This is, in addition to being a pure scientist, one has to be tactful and learn the art of getting the decisions implemented. This art can be achieved by experience as well as by getting training in social sciences, particularly psychology.

In fact, many managers may make a joke of OR as they think that the decisions made otherwise may be better. But being aware of its limitations, they need to be convinced of its utility, which doubtlessly forms the essential guideline for making better decisions.

EXERCISES

1. Discuss the origin and development of OR. What are the limitations of OR? How computer has helped in popularizing OR?

2. Describe the various objectives of OR.

3. What is OR? What are the characteristics and limitations of OR?

4. (a) what is the role of decision-making in OR? Define scientific decision-making and explain how it affects OR decisions.

   (b) Discuss the scope and limitation of OR.

5. Explain the role of computers in OR.

6. (a) Explain the role of OR in business.

   (b) Explain methodology of OR.

7. What are the limitations of using results from a mathematical model to make decision about operations?

8. Discuss briefly the scope of OR in financial management.

9. DefineOR. Give reasons why most definitions of OR are not satisfactory.

10. Explain the role OR in solving industrial problems.

UNIT-II

TRANSPOTATION PROBLEM

The most important and successful applications in the optimization refers to **transportation problem** (TP), that is a special class of the linear programming (LP) in the **operation research** (OR). Approach: The main objective of **transportation problem** solution methods is to minimize the cost or the time of **transportation**

The Transportation problem The general transportation problem is concerned with determining an optimal strategy for distributing a commodity from a group of supply centres,such as factories, called sources, to various receiving centers, such as warehouses, called destinations, in such a way as to minimise total distribution costs. Each source is able to supply a fixed number of units of the product, usually called the capacity or availability, and each destination has a fixed demand, often called the requirement.

Transportation models can also be used when a firm is trying to decide where to locate a new facility. Good financial decisions concerning facility location also attempt to minimize total transportation and production costs for the entire system. Setting up a Transportation problem To illustrate how to set up a transportation problem we consider the following example; Example 4.1 A concrete company transports concrete from three plants, 1, 2 and 3, to three construction sites, A, B and C. The plants are able to supply the following numbers of tons per week: Plant Supply (capacity) 1, 300 2, 300 3, 100 The requirements of the sites, in number of tons per week, are:

Construction site Demand (requirement) A, 200 B, 200 C, 300 The cost of transporting 1 ton of concrete from each plant to each site is shown in the figure 8 in Emalangeni per ton. For computational purposes it is convenient to put all the above information into a table, as in the simplex method. In this table each row represents a source and each column represents a destination. Sites P From PPPPPPPP To A B C Supply (Availability) 1 4 3 8 300 Plants 2 7 5 9 300 3 4 5 5 100 Demand (requirement) 200 200 300

Mathematical model of a transportation problem

Before we discuss the solution of transportation problems we will introduce the notation used to describe the transportation problem and show that it can be formulated as a linear programming problem. We use the following notation; xij = the number of units to be distributed from source i to destination j (i = 1, 2, . . . , m; j = 1, 2, . . . , n); si = supply from source i; dj = demand at destination j; cij = cost per unit distributed from source i to destination j With respect to Example 4.1 the decision variables xij are the numbers of tons transported from plant i (where i = 1, 2, 3) to each site j (where j = A, B, C) A basic assumption is that the distribution costs of units from source i to destination j is directly proportional to the number of units distributed. A typical cost and requirements table has the form shown on Table 4. Let Z be total distribution costs from all the m sources to the n destinations. In example 4.1 each term in the objective function Z represents the total cost of tonnage transported on one route. For example, in the route 2 −→ C, the term in 9x2C, that is: (Cost per ton = 9) × (number of tons transported = x2C)

Hence the objective function is: Z = 4x1A + 3x1B + 8x1C + 7x2A + 5x2B + 9x2C + 4x3A + 5x3B + 5x3C Notice that in this problem the total supply is 300 + 300 + 200 = 700 and the total demand is 200 + 200 + 300 = 700. Thus Total supply = total demand. In mathematical form this expressed as Xm i=1 si = Xn j=1 dj (47) This is called a balanced problem . In this unit our discussion shall be restricted to the balanced problems. In a balanced problem all the products that can be supplied are used to meet the demand. There are no slacks and so all constraints are equalities rather than inequalities as was the case in the previous unit. The formulation of this problem as a linear programming problem is presented as Minimise Z = Xm i=1 Xn j=1 cijxij, (48) subject to Xn j=1 xij = si , for i = 1, 2, . . . , m (49) Xn i=1 xij = dj , for j = 1, 2, . . . , n xij ≥ 0, for all i and j. Any linear programming problem that fits this special formulation is of the transportation type, regardless of its physical context. For many applications, the supply and demand quantities in the model will have integer values and implementation will require that the distribution quantities also be integers. Fortunately, the unit coefficients of the unknown variables in the constraints guarantee an optimal solution with only integer values.

It is assumed that the production costs per desk are identical at each factory. The only relevant costs are those of shipping from each source to each destination. The costs are shown in Table 5 P From PPPPPPPP To A B C D $5 $4 $3 E $8 $4 $3 F $9 $7 $5 Table 5: Transportation Costs per desk for Executive Furniture Corp. We proceed to construct a transportation table and label its various components as show in Table 6. We can now use the Northwest corner rule to find an initial feasible solution to the problem.

The Transportation and Assignment problems deal with assigning sources and jobs to destinations and machines. We will discuss the transportation problem first.

Suppose a company has m factories where it manufactures its product and n outlets from where the product is sold. Transporting the product from a factory to an outlet costs some money which depends on several factors and varies for each choice of factory and outlet. The total amount of the product a particular factory makes is fixed and so is the total amount a particular outlet can store. The problem is to decide how much of the product should be supplied from each factory to each outlet so that the total cost is minimum.

Let us consider an example.

Suppose an auto company has three plants in cities A, B and C and two major distribution centers in D and E. The capacities of the three plants during the next quarter are 1000, 1500 and 1200 cars. The quarterly demands of the two distribution centers are 2300 and 1400 cars. The transportation costs (which depend on the mileage, transport company etc) between the plants and the distribution centers is as follows:

Cost Table	Dist Center D	Dist Center E
Plant A	80	215
Plant B	100	108
Plant C	102	68

Which plant should supply how many cars to which outlet so that the total cost is minimum?

The problem can be formulated as a LP model:

Let be the amount of cars to be shipped from source i to destination j. Then our objective is to minimize the total cost which is The constraints are the ones imposed by the amount of cars to be transported from each plant and the amount each center can absorb.

The whole model is:

Minimize z =

subject to, and integer, i = 1,2,3, j = 1,2.

Vogel Approximation Method

Consider the transportation problem presented in the following table:

		Destination		
Origin	**1**	**2**	**3**	**Supply**
1	2	7	4	5
2	3	3	1	8
3	5	4	7	7
4	1	6	2	14
Demand	7	9	18	34

Solution.

Table 1

		Destination			
Origin	**1**	**2**	**3**	**Supply**	**Penalty**
1	5 2	7	4	~~5~~	2
2	3	3	1	8	2
3	5	4	7	7	1
4	1	6	2	14	1
Demand	~~7~~ 2	9	18	34	
Penalty	1	1	1		

The highest penalty occurs in the first row. The minimum cij in this row is c_{11} (i.e., 2). Hence, x_{11} = 5 and the first row is eliminated.

Now again calculate the penalty. The following table shows the computation of penalty for various rows and columns.

Final table

	Destination									
Origin	**1**	**2**	**3**	**Supply**	**Penalty**					
1	$\boxed{2}$ 5	7	4	~~5~~	2	-	-	-	-	-
2	3	$\boxed{3}$ 2	$\boxed{1}$ 6	~~8~~	2	2	2	2	3	3
3	5	$\boxed{4}$ 7	7	~~7~~	1	1	3	3	4	-
4	$\boxed{1}$ 2	6	$\boxed{2}$ 12	~~14~~	1	1	4	-	-	-
Demand	~~7~~	~~9~~	~~18~~	34						
Penalty	1	1	1							
	2	1	1							
	-	1	1							
	-	1	6							
	-	1	-							
	-	3	-							

EXERCISES

1. Explain steps involved in V.A.M.

2. Explain transportation problem giving examples

3. (a) How will you define transportation model? Explain its application.

(b) How will you define trans-shipment model? Explain its application.

(c) What is looping in transportation problem?

4. Write mathematical model for general transportation problem.

5. What do you understand by a balanced and an unbalanced transportation problem? How an unbalanced problem is tackled?

6. What are the conditions for the application of the optimality test in case of transportation problem? Briefly explain as to why these conditions should be satisfied?

7. Explain briefly the following:

(i) MODI method.

(ii) Loops in transportation problem.

8. (a) What do you understand from a transportation model of OR.? List various methods of solving a transportation problem. Which is the best method of solving transportation problem and why?

(b) ExplainVogel's approximation method.

(c) Explain any three methods of finding basic feasible solution of transportation problem.

9. Write a short note on dual of transportation model.

10.Explain sensitivity analysis in transportation problem when the changes take place in

(i) source capacities

(ii) destination requirements

(iii) unit transportation costs.

11. Explain any three methods of finding initial feasible solution of a transportation.

12. State the assignment model. Describe an algorithm for the solution of the solution of the assignment problem.

13. Write short note on the assignment problem and its applications.

14. Explain the following in the context of assignment problem:

(i) Balanced assignment problem

(ii)The Hungarian method

UNIT-III

Network analysis

Introduction

Network analysis is the general name given to certain specific techniques which can be used for the planning, management and control of projects. One definition of a project (from the Project Management Institute) is

A project is a temporary endeavour undertaken to create a "unique" product or service

This definition serves to highlight some essential features of a project

- it is temporary - it has a beginning and an end
- it is "unique" in some way

With regard to the use of the word unique I personally prefer to use the idea of "non-repetitive" or "non-routine", e.g. building the very first Boeing Jumbo jet was a project - building them now is a repetitive/routine manufacturing process, not a project.

We can think of many projects in real-life, e.g. building the Channel tunnel, building the London Eye, developing a new drug, etc

Typically all projects can be broken down into:

- separate *activities* (tasks/jobs) - where each activity has an associated duration or *completion time* (i.e. the time from the start of the activity to its finish)

- *precedence relationships* - which govern the order in which we may perform the activities, e.g. in a project concerned with building a house the activity "erect all four walls" must be finished before the activity "put roof on" can start

and the problem is to bring all these activities together in a coherent fashion to complete the project.

Two different techniques for network analysis were developed independently in the late 1950's - these were:

- PERT (for Program Evaluation and Review Technique); and
- CPM (for Critical Path Management).

PERT was developed to aid the US Navy in the planning and control of its Polaris missile program . This was a project to build a strategic weapons system, namely the first submarine launched intercontinental ballistic missile, at the time of the Cold War between the USA and Russia. Military doctrine at that time emphasised 'MAD - mutually assured destruction', namely if the other side struck first then sufficient nuclear weapons would remain to obliterate their homeland. That way peace was preserved. By the late 1950s the USA believed (or more importantly believed that the Russians believed) that American land based missiles and nuclear bombers were vulnerable to a first strike. Hence there was a strategic emphasis on completing thePolaris project as quickly as possible, cost was not an issue. However no one had ever build a submarine launched intercontinental ballistic missile before, so dealing with uncertainty was a key issue. PERT has the ability to cope with uncertain activitycompletion times (e.g. for a particular activity the most likely completion time is 4 weeks but it could be any time between 3 weeks and 8 weeks).

CPM was developed in the 1950's as a result of a joint effort by the DuPont Company and Remington Rand Univac. As these were commercial companies cost was an issue, unlike the Polaris project mentioned above. In CPM the emphasis is on the trade-off between the cost of the project and its overall completion time (e.g. for certain activities it may be possible to decrease their completion times by spending more money - how does this affect the overall completion time of the project?)

Modern commercial software packages tend to blur the distinction between PERT and CPM and include options for uncertain activity completion times and project completion time/project cost trade-off analysis. Note here that many such packages exist for doing network analysis.

There is no clear terminology in the literature and you will see this area referred to by the phrases: network analysis, PERT, CPM, PERT/CPM, critical path analysis and project planning.

Network analysis is a vital technique in **PROJECT MANAGEMENT**. It enables us to take a ***systematic quantitative structured approach*** to the problem of managing a project through to successful completion. Moreover, as will become clear below, it has a graphical representation which means it can be understood and used by those with a less technical background.

5 Important Objectives of Network Analysis

(a) To minimize idle resources:

Allowing for large variations in the use of limited resources may disturb the whole plan. Thus, efforts should be made to avoid the cost incurred due to idle resources.

(b) To minimize the total project cost:

The total cost of the project can be calculated and then efforts can be made to minimize the total cost by calculating the cost of delay in the completion of an activity of the project in addition to the cost of the resources required to carry out the jobs at various speeds (i.e. normal or over time rates of pay).

(c) To trade off between time and cost of project:

The idea of trade off between time and cost of project is centred on the idea that duration of same activities can be cut down if additional resources are allocated to them.

For technical reasons, the durations may not be reduced indefinitely. Similarly, there is also a most cost efficient duration called 'normal point' & stretching the activity beyond it may lead to [a] rise in direct costs.

(d) To minimize production delays, interruption and conflict:

This is achieved by identifying all activities involved in the project, their precedence constraints, etc.

(e) To minimize idle resources:

Allowing for large variations in the use of limited resources may disturb the whole plan. Thus, efforts should be made to avoid the cost incurred due to resources.

PERT/CPM for Project Scheduling & Management

INTRODUCTION

Basically, CPM (Critical Path Method) and PERT (Programme Evaluation Review Technique) are project management techniques, which have been created out of the need of Western industrial and military establishments to plan, schedule and control complex projects.

Brief History of CPM/PERT

CPM/PERT or Network Analysis as the technique is sometimes called, developed along two parallel streams, one industrial and the other military.

CPM was the discovery of M.R.Walker of E.I.Du Pont de Nemours & Co. and J.E.Kelly of Remington Rand, circa 1957. The computation was designed for the UNIVAC-I computer. The first test was made in 1958, when CPM was applied to the construction of a new chemical plant. In March 1959, the method was applied to a maintenance shut-down at the Du Pont works in Louisville, Kentucky. Unproductive time was reduced from 125 to 93 hours.

PERT was devised in 1958 for the POLARIS missile program by the Program Evaluation Branch of the Special Projects office of the U.S.Navy, helped by the Lockheed Missile Systems division and the Consultant firm of Booz-Allen & Hamilton. The calculations were so arranged so that they could be carried out on the IBM Naval Ordinance Research Computer (NORC) at Dahlgren, Virginia.

1.2 Planning, Scheduling & Control

Planning, Scheduling (or organising) and Control are considered to be basic Managerial functions, and CPM/PERT has been rightfully accorded due importance in the literature on Operations Research and Quantitative Analysis.

Far more than the technical benefits, it was found that PERT/CPM provided a focus around which managers could brain-storm and put their ideas together. It proved to be a great communication medium by which thinkers and planners at one level could communicate their ideas, their doubts and fears to another level. Most important, it became a useful tool for evaluating the performance of individuals and teams.

There are many variations of CPM/PERT which have been useful in planning costs, scheduling manpower and machine time. CPM/PERT can answer the following important questions:

How long will the entire project take to be completed? What are the risks involved?

Which are the critical activities or tasks in the project which could delay the entire project if they were not completed on time?

Is the project on schedule, behind schedule or ahead of schedule?

If the project has to be finished earlier than planned, what is the best way to do this at the least cost?

The Framework for PERT and CPM

Essentially, there are six steps which are common to both the techniques. The procedure is listed below:

I. Define the Project and all of it's significant activities or tasks. The Project (made up of several tasks) should have only a single start activity and a single finish activity.
II. Develop the relationships among the activities. Decide which activities must precede and which must follow others.
III. Draw the "Network" connecting all the activities. Each Activity should have unique event numbers. Dummy arrows are used where required to avoid giving the same numbering to two activities.
IV. Assign time and/or cost estimates to each activity
V. Compute the longest time path through the network. This is called the critical path.
VI. Use the Network to help plan, schedule, monitor and control the project.

The Key Concept used by CPM/PERT is that a small set of activities, which make up the longest path through the activity network control the entire project. If these "critical" activities could be identified and assigned to responsible persons, management resources could be optimally used by concentrating on the few activities which determine the fate of the entire project.

Non-critical activities can be replanned, rescheduled and resources for them can be reallocated flexibly, without affecting the whole project.

Five useful questions to ask when preparing an activity network are:

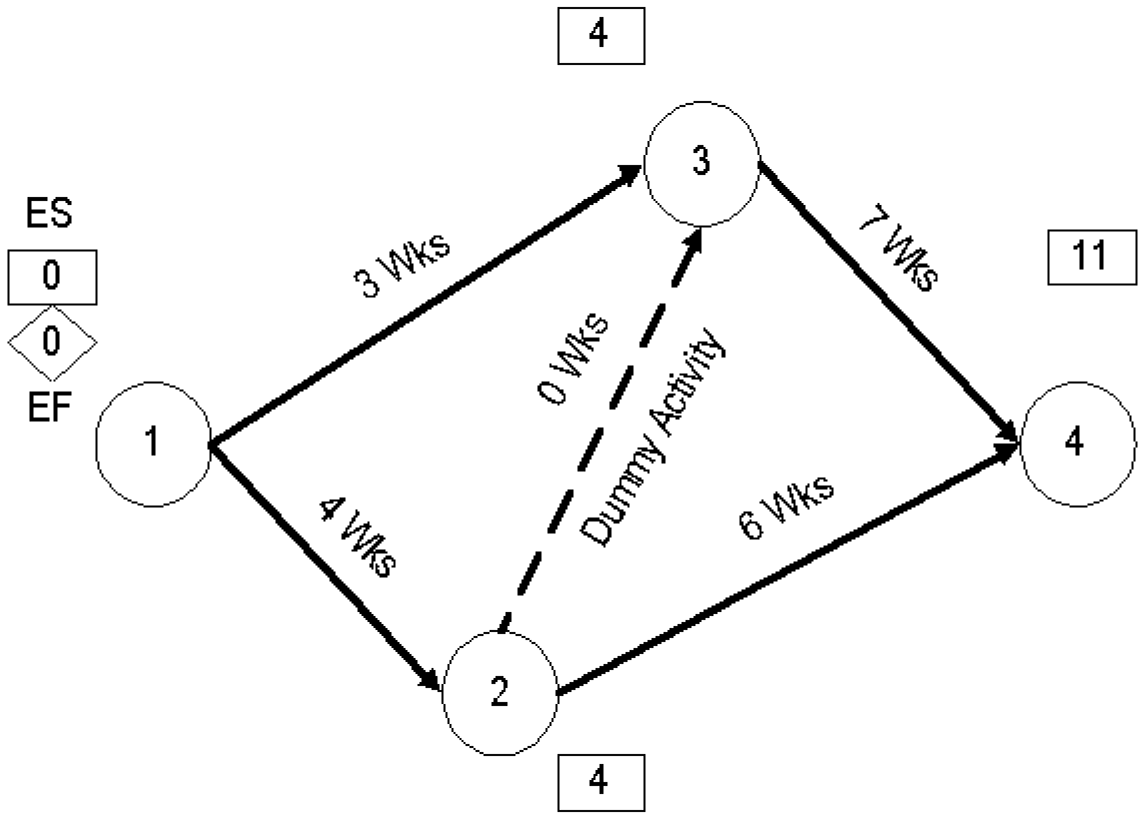
- Is this a Start Activity?
- Is this a Finish Activity?
- What Activity Precedes this?
- What Activity Follows this?
- What Activity is Concurrent with this?

Some activities are serially linked. The second activity can begin only after the first activity is completed. In certain cases, the activities are concurrent, because they are independent of each other and can start simultaneously. This is especially the case in organisations which have supervisoryresources so that work can be delegated to various departments which will be responsible for the activities and their completion as planned.

When work is delegated like this, the need for constant feedback and co-ordination becomes an important senior management pre-occupation.

1.4 Drawing the CPM/PERT Network

Each activity (or sub-project) in a PERT/CPM Network is represented by an arrow symbol. Each activity is preceded and succeeded by an event, represented as a circle and numbered.

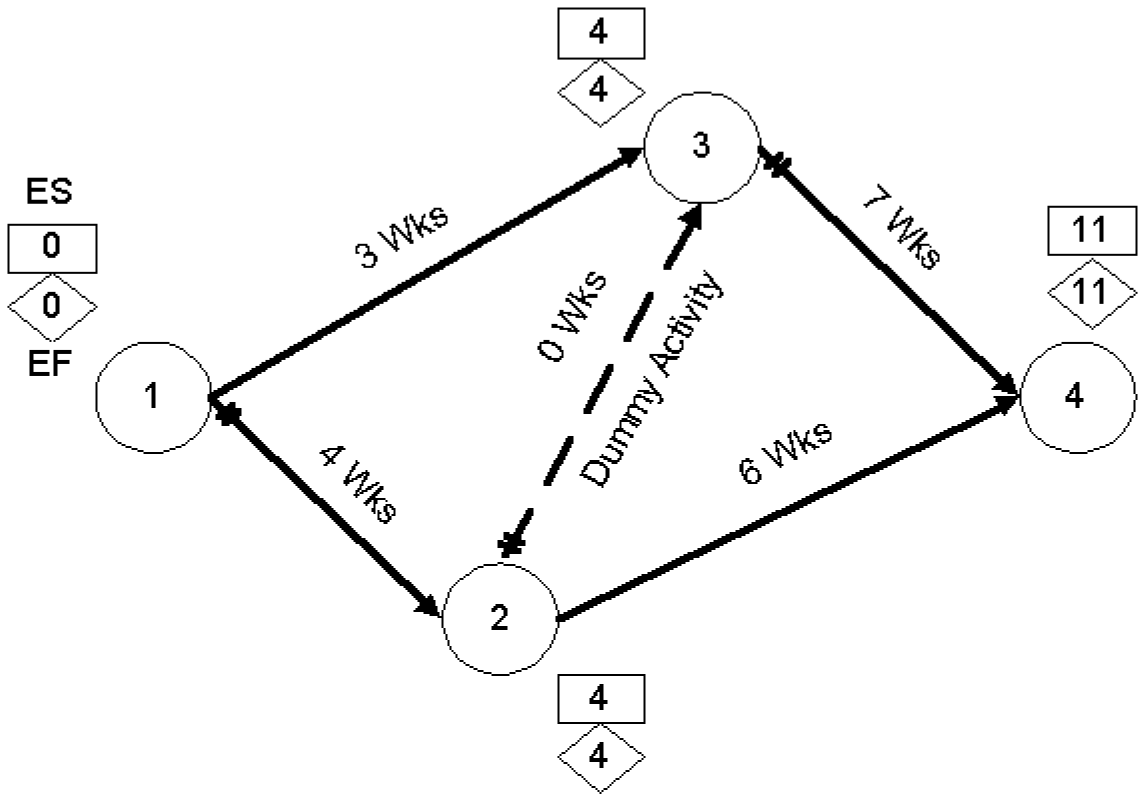


At Event 3, we have to evaluate two predecessor activities – Activity 1-3 and Activity 2-3, both of which are predecessor activities. Activity 1-3 gives us an Earliest Start of 3 weeks at Event 3. However, Activity 2-3 also has to be completed before Event 3 can begin. Along this route, the Earliest Start would be 4+0=4. The rule is to take the longer (bigger) of the two Earliest Starts. So the Earliest Start at event 3 is 4.

Similarly, at Event 4, we find we have to evaluate two predecessor activities – Activity 2-4 and Activity 3-4. Along Activity 2-4, the Earliest Start at Event 4 would be 10 wks, but along Activity 3-4, the EarliestStart at Event 4 would be 11 wks. Since 11 wks is larger than 10 wks, we select it as the Earliest Start at Event 4.**We have now found the longest path through the network. It will take 11 weeks along activities 1-2, 2-3 and 3-4. This is the Critical Path.**

The Backward Pass – Latest Finish Time Rule

To make the Backward Pass, we begin at the sink or the final event and work backwards to the first event.



At Event 3 there is only one activity, Activity 3-4 in the backward pass, and we find that the value is 11-7 = 4 weeks. However at Event 2 we have to evaluate 2 activities, 2-3 and 2-4. We find that the backward pass through 2-4 gives us a value of 11-6 = 5 while 2-3 gives us 4-0 = 4. We take the **smaller value** of 4 on the backward pass.

Tabulation & Analysis of Activities

We are now ready to tabulate the various events and calculate the Earliest and Latest Start and Finish times. We are also now ready to compute the SLACK or TOTAL FLOAT, which is defined as the difference between the Latest Start and Earliest Start.

Event	Duration(Weeks)	Earliest Start	Earliest Finish	Latest Start	Latest Finish	Total Float
1-2	4	0	4	0	4	0
2-3	0	4	4	4	4	0
3-4	7	4	11	4	11	0
1-3	3	0	3	1	4	1
2-4	6	4	10	5	11	1

- The Earliest Start is the value in the rectangle near the tail of each activity
- The Earliest Finish is = Earliest Start + Duration
- The Latest Finish is the value in the diamond at the head of each activity
- The Latest Start is = Latest Finish – Duration

There are two important types of Float or Slack. These are Total Float and Free Float.

TOTAL FLOAT is the spare time available when all preceding activities occur at the **earliest** possible times and all succeeding activities occur at the **latest** possible times.

- Total Float = Latest Start – Earliest Start

Activities with zero Total float are on the Critical Path

Inventory theory

(or more formally the mathematical theory of inventory and production) is the sub-specialty within operations research and operations management that is concerned with the design of production/inventory systems to minimize costs: it studies the decisions faced by firms and the military in connection with manufacturing, warehousing, supply chains, spare part allocation and so on and provides the mathematical foundation for logistics. The **inventory control problem** is the problem faced by a firm that must decide how much to order in each time period to meet demand for its products. The problem can be modeled using mathematical techniques of optimal control, dynamic programming and network optimization. The study of such models is part of inventory theory.

Inventory models

The mathematical approach is typically formulated as follows: a store has, at time , items

in stock. It then orders (and receives) items, and sells items, where follows a given probability distribution. Thus:

Whether is allowed to go negative, corresponding to back-ordered items, will depend on the specific situation; if allowed there will usually be a penalty for back orders. The store has costs that are related to the number of items in store and the number of items ordered:

Often this will be in additive form:

The store wants to select in an optimal way, i.e. to minimize

Many other features can be added to the model, including multiple products

(denoted), upper bounds on inventory and so on. Inventory models can be based on different assumptions:[1][2]

- Nature of demand: constant, deterministically time-varying or stochastic
- Costs: variable versus fixed
- Flow of time: discrete versus continuous
- Lead Time: deterministic or stochastic
- Time horizon: finite versus infinite (T=+∞)
- Presence or absence of back-ordering
- Production rate: infinite, deterministic or random
- Presence or absence of quantity discounts
- Imperfect quality
- Capacity: infinite or limited
- Products: one or many
- Location: one or many
- Echelons: one or many

Classic models

Although the number of models described in the literature is immense, the following is a list of classics:

- Infinite fill rate for the part being produced: Economic order quantity model, a.k.a. Wilson EOQ Model
- Constant fill rate for the part being produced: Economic production quantity model
- Demand is random, only one replenishment: classical Newsvendor model
- Demand is random, continuous replenishment: Base Stock Model
- Demand varies deterministically over time: Dynamic lot size model or Wagner-Whitin model
- Demand varies deterministically over time: Silver–Meal heuristic
- Several products produced on the same machine:

Deterministic and Probabilistic models and thinking

The way we understand and make sense of variation in the world affects decisions we make.

Part of understanding variation is understanding the difference between deterministic and probabilistic (stochastic) models. The NZ curriculum specifies the following learning outcome: "Selects and uses appropriate methods to investigate probability situations including experiments, simulations, and theoretical probability, **distinguishing between deterministic and probabilistic models.**" This is at level 8 of the curriculum, the highest level of secondary schooling. Deterministic and probabilistic models are not familiar to all teachers of mathematics and statistics, so I'm writing about it today.

Model

The term, model, is itself challenging. There are many ways to use the word, two of which are particularly relevant for this discussion. The first meaning is "mathematical model, as a decision-making tool". This is the one I am familiar with from years of teaching Operations Research. The second way is "way of thinking or representing an idea". Or something like that. It seems to come from psychology.

When teaching mathematical models in entry level operations research/management science we would spend some time clarifying what **we** mean by a model. I have written about this in the post, "All models are wrong."

In a simple, concrete incarnation, a model is a representation of another object. A simple example is that of a model car or a Lego model of a house. There are aspects of the model that are the same as the original, such as the shape and ability to move or not. But many aspects of the real-life object are missing in the model. The car does not have an internal combustion engine, and the house has no soft-furnishings. (And very bumpy floors). There is little purpose for either of these models, except entertainment and the joy of creation or ownership.

Deterministic models

A deterministic model assumes certainty in all aspects. Examples of deterministic models are timetables, pricing structures, a linear programming model, the economic order quantity model, maps, accounting.

Probabilistic or stochastic models

Most models really should be stochastic or probabilistic rather than deterministic, but this is often too complicated to implement. Representing uncertainty is fraught. Some more common stochastic models are queueing models, markov chains, and most simulations.

For example when planning a school formal, there are some elements of the model that are deterministic and some that are probabilistic. The cost to hire the venue is deterministic, but the number of students who will come is probabilistic. A GPS unit uses a deterministic model to decide on the most suitable route and gives a predicted arrival time. However we know that the actual arrival time is contingent upon all sorts of aspects including road, driver, traffic and weather conditions.

Purchasing model with shortages

In this model, shortages are allowed and consequently a shortage cost is incurred. Let the short ages be
denoted by `S' for every cycle and shortage cost by $C4$ per item per unit time.

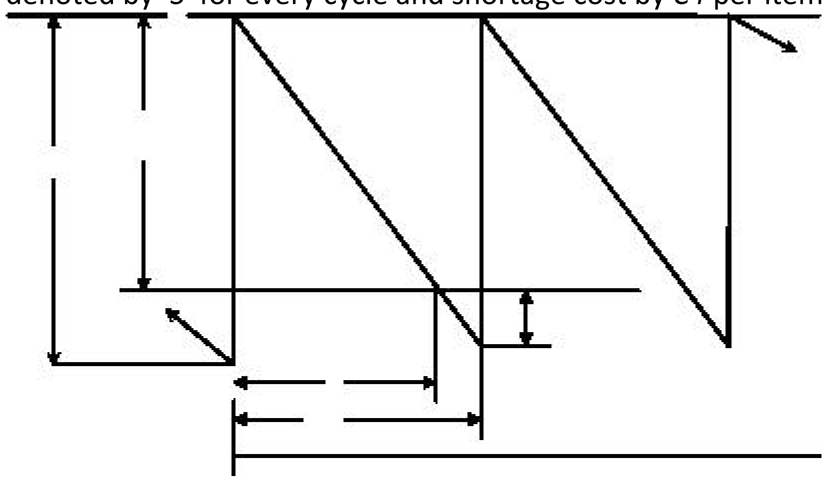
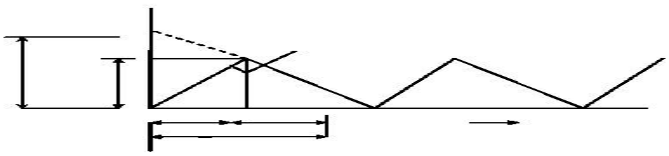


Fig. 3 shows that the back ordering is possible (i.e.) once an order is received, any shortages can be made
up
as the items are received. Consequently shortage costs are due to being short of stock for a peri od of time.
The cost per period includes four cost components.
Total cost per period = Item cost + Order cost + Holding cost + Shortage cost
Item cost per period = (item cost) × (number of items/period)
= $C1Q$
(13)
Order cost per period = $C2$
Let $t1$ be the time period during which only the items are held in stock. Let the maximum invent ory be denoted by Im
and this is equal to (Q S) or Im = (Q S)

Manufacturing model with no shortages

In this model the following assumptions are made:
(1) Demand is at a constant rate (D).
(2)All cost coefficients ($C1$, $C2$, $C3$) are constants.

(3) There is no shortage cost, or $C4$ = 0.
(4)The replacement rate is finite and greater than the demand rate. This is also called replenish ment rate
or manufacturing rate, denoted by R



The total cost of inventory per period is the sum of three components: item cost, order cost and items holding cost.
Let Im be the maximum inventory, $t1$ be the time of manufacture and $t2$ be the time during whi ch there is no supply.
In this model, all items required for a cycle are not stored at the beginning as
in Wilson's Model. The items are
manufactured at a higher rate than the demand so that the difference (RD)
is the existing inventory till the items are
exhausted.
Item cost/period = $C1Q$
(32)
Order cost/period = $C2$

ORDER QUANTITY WITH PRICE-BREAK

The concept of Economic Order Quantity fails in certain cases where there is
a discount offered when purchases are
made in large quantities. Certain manufacturers offer reduced rate for items when a larger quant ity is ordered. It may
appear that the inventory holding cost may increase if large quantities of items are ordered. But i f the discount offered
is
so attractive that it even outweighs the holding cost, the probably the order at levels other than the EOQ would be
economical. An illustration is given in the following example and the rationale is explained.
Example: A company uses 12000 items per year supplied ordinarily at
a price of Rs. 3.00 per item. Carrying costs
are 16% of the value of the average inventory and the ordering costs are Rs. 20 per order.
Order size
Price per item

Less than 2000
Rs. 3.00
2000 to 3999
Rs. 290
4000 or more
Rs. 2.85
Compute the economic order size.
2 ×12000 × 20
EOQ =
= 1000
3.00 × 0.16
EOQ at Rs. 2.90 per item
2 ×12000 × 20
= 1017
2.90 × 0.16
EOQ at Rs. 2.85 per item
2 ×12000 × 20
= 1026
2.85 × 0.16
The EOQ at Rs. 2.90 per item = 1017. But the price per item is Rs. 2.9 only if the items are ordere d in the range of2000 to 3999. This is therefore an infeasible solution. Similarly the EOQ at Rs. 2. 85 per item is 1026. This price is

valid only for items ordered in the range 4000 or more. This is also an infeasible solution.

PRICE

A number of studies in the economics and marketing literature have suggested quantity discount as a mechanism to achieve co-ordination between the seller (or the manufacturer) and the buyer (or the retailer). The basic aim of co-ordination is that a quantity discount schedule can be designed such that the objectives of the seller and buyer are incentive compatible with the maximum channel gain. Quantity discount has been proposed in two separate research streams as a tool for achieving incentive compatible co-ordination within the chain. Suppose the supply chain consists of two members: The seller who determines the wholesale price and the buyer who chooses his optimal order quantity and the retail price. In the management literature, the channels' total transaction cost can be minimized by properly designing the quantity discount schedule so that the buyer orders the channels' economical order quantity. On the other hand, the supplier can offer the buyer a quantity discount, which induces the buyer to choose his price at the joint optimum level. This eliminates double marginalization, and increases market demand due to the lower retail price, which benefits the whole system. However, other purpose of quantity discount is found in literature, which can be summarized as: 1. Perfect price discrimination against a single customer or a set of homogeneous customers. *Corresponding author. E-mail: lahjiaziz@yahoo.com. 2. Partial price discrimination against a set of heterogeneous customers. Dolan (1987) investigated the effect of quantity discount on joint economic lot size models, while Monahan (1985) developed the model for the quantity discount, from the perspective of the seller with constraints imposed to ensure sufficient benefit to the buyer. In this chapter, a joint economic lot-size model is developed under Fuzzy environment. Optimum order quantity and optimum number of price breaks are obtained through the techniques of Fuzzy mathematics. Cost function of both buyer and seller is represented by Fuzzy membership function and model is solved for optimum results by reformulating it as Fuzzy linear programming problem. ASSUMPTIONS AND NOTATIONS Assumptions Following are the assumptions made for developing the model under study: 1. The simpler EOQ model with deterministic demand, no stock outs, no backlogs and deterministic lead-time, can describe the buyer's inventory policy. 2. The seller has the knowledge of the holding and ordering costs governing the buyer's inventory policy. 3. There is no competitive price reaction to the seller's discount policy. 4. The demand is fixed at uniform rate 5. There is a single buyer. THE BUYER'S PROBLEM When there is no channel of co-ordination, with the knowledge of buyer's reaction as a function of the seller's decision variable, the seller can optimize his own profit. The buyer's problem is to choose an order quantity that minimizes the total cost consisting of product cost, set-up cost and inventory cost. [Dada et. el.(1987)]: Q H D Q S D C (Q) pD b b bb = + + ... (1) Differentiating Equation (1) w.r.t. Q and equating it to zero, we get the buyer's optimal order quantity (EOQ) as: Q *= b b H 2S D ... (2) This is independent of the purchasing cost. The total cost for the buyer is: b * pD 2DSbHb C (Q) = + ... (3) It will be noted that in this formulation, the inventory holding cost is assumed to be per unit cost. Thus, this EOQ does not depend on the wholesale price. When the inventory cost depends upon the value of the wholesale price p, then EOQ will be: Q *= H p 2 S D b b ... (4) THE SELLER'S PROBLEM The seller's problem is to maximize his profit by influencing the buyer's order quantity. The seller can design the wholesale price as a function of the order quantity thereby providing an incentive to the buyer to choose a different order quantity. In the cost minimization problem, the seller's total cost is given by:) 2 Q) H (Q D C (Q) S (s = s − s ... (5) C (Q*) s = ( ) b s s b H bS b (H S − H S) D / 2 ... (6)

with the resulting joint channel cost: J(Q*) C (Q*) C (Q*) = b + s = b b s s b HbSbpD+(H (2S +S)−H S) D/2 ... (7) On the other hand, when the channel members' coordinate to minimizes the joint cost we have: 2 Q (S S) (H H) Q D J(Q) = pD+ b + s + b − s ... (8) The jointly optimal order quantity can be derived from Equation (7) as: b s b s H H 2(S S)D Q** − + = ... (9) The resulting joint channel cost at Q ** is given as: J(Q**) pD 2(D(H H)(S S) = + b − s b + s ... (10) since S 0, H 0 s > s > and Q **>Q *. This implies that the buyer's cost given in Equation (1) increases as its order quantity deviates from Q *. Also by definition, J(Q **) < J(Q *). Therefore: Cb (Q **) + Cs (Q **) Cb (Q *) +Cs (Q *), Which implies that: Cb (Q **)-Cb (Q *) Cs (Q *)-Cs (Q **). That is, when the channel members coordinate to minimize the joint cost, the decrease in the seller's cost is more than the increase in the buyer's cost when the order quantity is increased from Q * to Q **. This provides the seller a supply that can be used to motivate the buyer to order a higher quantity by passing some of these savings to cover the extra buyer's cost. Given the buyer's optimal order quantity policy, the seller's problem is to maximize his total profit by adjusting the wholesale price downward in order to motivate the buyer to change the order quantity for the benefit of the whole channel.

EXERCISES

1. Discuss the similarities and differences of CPM and PERT.

2. Discusss applications of PERT/CPM in project planning and explain the difference between them.

3. State the requirements for the application of PERT technique and practical limitations of using PERT. How does PERT differ from CPM?

4. Write a detailed note on the applications of network techniques.

5. Answer the following:
   i) Advantages of Network Models.
   ii) Difficulties in using Network Models
   iii) Application of Network Techniques.
iv) Compare and contrast CPM and PERT Models

UNIT- IV

DECISION THEORY

Decision theory (or the **theory of choice**) is the study of the reasoning underlying an agent's choices.[1] Decision theory can be broken into two branches: normative decision theory, which gives advice on how to make the best decisions, given a set of uncertain beliefs and a set of values; and descriptive decision theory, which analyzes how existing, possibly irrational agents actually make decisions.

Closely related to the field of game theory,[2] decision theory is concerned with the choices of individual agents whereas game theory is concerned with interactions of agents whose decisions affect each other. Decision theory is an interdisciplinary topic, studied by economists, statisticians, psychologists, biologists,[3] political and other social scientists, philosophers,[4] and computer scientists.

Empirical applications of this rich theory are usually done with the help of statistical and econometric methods, especially via the so-called choice models, such as probit and logitmodels. Estimation of such models is usually done via parametric, semi-parametric and non-parametric maximum likelihood methods.[5]

Choice under uncertainty

The area of choice under uncertainty represents the heart of decision theory. Known from the 17th century (Blaise Pascal invoked it in his famous wager, which is contained in his *Pensées*, published in 1670), the idea of expected value is that, when faced with a number of actions, each of which could give rise to more than one possible outcome with different probabilities, the rational procedure is to identify all possible outcomes, determine their values (positive or negative) and the probabilities that will result from each course of action, and multiply the two to give an "expected value", or the average expectation for an outcome; the action to be chosen should be the one that gives rise to the highest total expected value. In 1738, Daniel Bernoulli published an influential paper entitled *Exposition of a New Theory on the Measurement of Risk*, in which he uses the St. Petersburg paradox to show that expected value theory must be normatively wrong. He gives an example in which a Dutch merchant is trying to decide whether to insure a cargo being sent from Amsterdam to St Petersburg in winter. In his solution, he defines a utility function and computes expected utility rather than expected financial value (see[7] for a review).

In the 20th century, interest was reignited by Abraham Wald's 1939 paper[8] pointing out that the two central procedures of sampling-distribution-based statistical-theory, namely hypothesis testing and parameter estimation, are special cases of the general decision problem. Wald's paper renewed and synthesized many concepts of statistical theory, including loss functions, risk functions, admissible decision rules, antecedent distributions, Bayesian procedures, and minimax procedures. The phrase "decision theory" itself was used in 1950 by E. L. Lehmann.[9]

The revival of subjective probability theory, from the work of Frank Ramsey, Bruno de Finetti, Leonard Savage and others, extended the scope of expected utility theory to situations where subjective probabilities can be used. At the time, von Neumann and Morgenstern theory of expected utility[10] proved that expected utility maximization followed from basic postulates about rational behavior.

Daniel Kahneman

The work of Maurice Allais and Daniel Ellsberg showed that human behavior has systematic and sometimes important departures from expected-utility maximization. The prospect theory of Daniel Kahneman and Amos Tversky renewed the empirical study of economic behavior with less emphasis on rationality presuppositions. Kahneman and Tversky found three regularities – in actual human decision-making, "losses loom larger than gains"; persons focus more on *changes* in their utility-states than they focus on absolute utilities; and the estimation of subjective probabilities is severely biased by anchoring.

WHAT KINDS OF DECISION NEED A THEORY:

Intertemporal choice is concerned with the kind of choice where different actions lead to outcomes that are realised at different points in time. If someone received a windfall of several thousand dollars, they could spend it on an expensive holiday, giving them immediate pleasure, or they could invest it in a pension scheme, giving them an income at some time in the future. What is the optimal thing to do? The answer depends partly on factors such as the expected rates of interest and inflation, the person's life expectancy, and their confidence in the pensions industry. However even with all those factors taken into account, human behavior again deviates greatly from the predictions of prescriptive decision theory, leading to alternative models in which, for example, objective interest rates are replaced by subjective discount rates.

Interaction of decision makers

Some decisions are difficult because of the need to take into account how other people in the situation will respond to the decision that is taken. The analysis of such social decisions is more often treated under the label of game theory, rather than decision theory, though it involves the same mathematical methods. From the standpoint of game theory most of the problems treated in decision theory are one-player games (or the one player is viewed as playing against an impersonal background situation). In the emerging socio-cognitiveengineering, the research is especially focused on the different types of distributed decision-making in human organizations, in normal and abnormal/emergency/crisis situations.[11]

Complex decisions

Other areas of decision theory are concerned with decisions that are difficult simply because of their complexity, or the complexity of the organization that has to make them. Individuals making decisions may be limited in resources or are boundedly rational (have finite time or intelligence); in such cases the issue, more than the deviation between real and optimal behaviour, is the difficulty of determining the optimal behaviour in the first place. One example is the model of economic growth and resource usage developed by the Club of Rome to help politicians make real-life decisions in complex situations[citation needed]. Decisions are also affected by whether options are framed together or separately; this is known as the distinction bias.

SHORTCOMINGS OF EXPECTED MONETARY VALUE, UTILITY

In this section, we examine non-recurring decision problems in which the consequences cannot be measured in monetary terms. In theory, at least, these problems may be resolved by establishing the ìutilitiesî of the consequences, subjectively estimating the probabilities of the possible events, and selecting the act with the highest expected utility. Two examples will illustrate the nature of the problem and the method of resolution. Example 3.2 You are considering buying a ticket for a certain lottery. The ticket costs $100 and the lottery will be conducted only once. This is a rather crude lottery: a coin will be tossed; if it turns up heads, you will receive $250; if it turns up tails, you will get nothing. Should you buy this ticket or not? Alternatives (Acts) Events Probability Buy Do not buy Heads 0.5 +150 0 Tails 0.5 −100 0 Expected profit: 25 0 The consequences shown in the payoff table represent profit, i.e., the difference between revenue and cost. If you buy the ticket and the coin turns up heads, the profit is $250 − $100, or $150; if you buy and tails shows up, your profit if $0 − $100, or −$100. If you do not buy the ticket, the profit is $0, regardless of what the outcome of the toss might have been. Assuming that the coin is fair, the probabilities of heads and tails are 0.5 each. The expected profit of the act Buy is $25, while that of Do Not Buy is $0. The first act has the highest expected profit, and, according to this criterion, is the best act.

Decision-making under Certainty:

A condition of certainty exists when the decision-maker knows with reasonable certainty what the alternatives are, what conditions are associated with each alternative, and the outcome of each alternative. Under conditions of certainty, accurate, measurable, and reliable information on which to base decisions is available.

The cause and effect relationships are known and the future is highly predictable under conditions of certainty. Such conditions exist in case of routine and repetitive decisions concerning the day-to-day operations of the business.

Decision-making under Risk:

When a manager lacks perfect information or whenever an information asymmetry exists, risk arises. Under a state of risk, the decision maker has incomplete information about available alternatives but has a good idea of the probability of outcomes for each alternative.

While making decisions under a state of risk, managers must determine the probability associated with each alternative on the basis of the available information and his experience.

Modern Approaches to Decision-making under Uncertainty:

There are several modern techniques to improve the quality of decision-making under conditions of uncertainty.

The most important among these are:

(1) Risk analysis,

(2) Decision trees and

(3) preference theory.

Risk Analysis:

Managers who follow this approach analyze the size and nature of the risk involved in choosing a particular course of action.

For instance, while launching a new product, a manager has to carefully analyze each of the following variables the cost of launching the product, its production cost, the capital investment required, the price that can be set for the product, the potential market size and what percent of the total market it will represent.

Risk analysis involves quantitative and qualitative risk assessment, risk management and risk communication and provides managers with a better understanding of the risk and the benefits associated with a proposed course of action. The decision represents a trade-off between the risks and the benefits associated with a particular course of action under conditions of uncertainty.

Decision Trees:

These are considered to be one of the best ways to analyze a decision. A decision-tree approach involves a graphic representation of alternative courses of action and the possible outcomes and risks associated with each action.

By means of a **"tree"** diagram depicting the decision points, chance events and probabilities involved in various courses of action, this technique of decision-making allows the decision-maker to trace the optimum path or course of action.

Preference or Utility Theory:

This is another approach to decision-making under conditions of uncertainty. This approach is based on the notion that individual attitudes towards risk vary. Some individuals are willing to take only smaller risks ("risk averters"), while others are willing to take greater risks ("gamblers"). Statistical probabilities associated with the various courses of action are based on the assumption that decision-makers will follow them.

For instance, if there were a 60 percent chance of a decision being right, it might seem reasonable that a person would take the risk. This may not be necessarily true as the individual might not wish to take the risk, since the chances of the decision being wrong are 40 percent. The attitudes towards risk vary with events, with people and positions.

Top-level managers usually take the largest amount of risk. However, the same managers who make a decision that risks millions of rupees of the company in a given program with a 75 percent chance of success are not likely to do the same with their own money.

Moreover, a manager willing to take a 75 percent risk in one situation may not be willing to do so in another. Similarly, a top executive might launch an advertising campaign having a 70 percent chance of success but might decide against investing in plant and machinery unless it involves a higher probability of success.

Though personal attitudes towards risk vary, two things are certain.

Firstly, attitudes towards risk vary with situations, i.e. some people are risk averters in some situations and gamblers in others.

Secondly, some people have a high aversion to risk, while others have a low aversion.

Most managers prefer to be risk averters to a certain extent, and may thus also forego opportunities. When the stakes are high, most managers tend to be risk averters; when the stakes are small, they tend to be gamblers.

THE THEORY OF GAMES:

The theory of games (or game theory or competitive strategies) is a mathematical theory that deals with the general features of competitive situations. *This theory is helpful when two or more individuals or organisations with conflicting objectives try to make decisions.* In such situations, a decision made by one decision-maker affects the decision made by one or more of the remaining decision-makers and the final outcome depends upon the decision of all the parties. Such situations often arise in the fields of business, industry, economics, sociology and military training. This theory is applicable to a wide variety of situations such as two players struggling to win at chess, candidates fighting an election, two enemies planning war tactics, firms struggling to maintain their market shares, launching advertisement campaigns by companies marketing competing product, negotiations between organisations and unions, etc. These situations differ from the ones we have discussed so far wherein nature was viewed as a *harmless opponent*.

The theory of games is based on the *minimax principle* put forward by J. von Neumann which implies that each competitor will act so as to minimize his maximum loss (or maximize his minimum gain) or achieve *best of the worst*. So far only simple competitive problems have been analysed by this mathematical theory. The theory does not describe how a game should be played; it describes only the procedure and principles by which plays should be selected.

CHARACTERISTICS OF GAMES

A competitive game has the following characteristics

(i) There are *finite* number of participants or competitors. If the number of participating is 2, the game is called two-person game; for number greater than two, it is called n-person game.

(ii) Each participant has available to him a list of *finite* number of possible courses of action. The list may not be same for each participant.

(iii) Each participant knows all the *possible choices* available to others but does not know which of them is going to be chosen by them.

(iv) A*play* is said to occur when each of the participants chooses one the courses of action available to him. The choices are assumed to be made simultaneously so that no participant knows the choices made by others until he has decided his own.

(v) Every combination of courses of action determines an outcome which results in gains to the participants. The gain may be positive, negative or zero. Negative gain is called a loss.

(vi) The gain of a participant depends not only on his own actions but also those of others.

(vii) The gains (payoffs) for each and every play are fixed and specified in advance and are known to each player. Thus each player knows fully the information contained in the payoff matrix.

(viii) The players make individual decisions without direct communication.

Mixed strategies

Each player, instead of selecting pure strategies only, may play all his strategies according to a predetermined set of ratios. Let x_i , $i = 1, 2, \cdots, m$ and y_j , $j = 1, 2, \cdots, n$ be the row and column ratios representing the relative frequencies by which I and II, respectively, select their pure strategies. Then $x_i \geq 0$,$y_j \geq 0$ and Xm i=1 x_i = Xn j=1 y_j = 1 . II y_1 y_2 $\cdots$ y_n x_1 a_{11} a_{12} $\cdots$ a_{1n} I . . . . x_m a_{m1} a_{m2} $\cdots$ a_{mn} We can think of x_i and y_j as probabilities (generated by some random mechanism) by which I and II select their i th and j th pure strategies, respectively. The solution of the mixed strategy problem is based also on the minimax criterion. The only difference is that I selects the ratios xi (instead of the pure strategies i) which maximize the minimum expected payoff in a column, while II selects the ratios yj (instead of pure strategies j) which minimize the maximum expected payoff in a row. Mathematically, player I selects xi(xi ≥ 0, Pxi = 1) which will yield Max xi (Min·Xm i=1 ai1xi , Xm i=1 ai2xi , · · · , Xm i=1 ainxn ¸) , and player II selects yj (yj ≥ 0, Pyj = 1) which will yield Min yj ⍰ ⍰ ⍰ Max·Xn j=1 a1jyj , Xn j=1 a2jyj , · · · , Xn j=1 amjyj ¸ ⍰ ⍰ ⍰ . These values are referred to as the maximin and minimax expected payoffs, respectively. As in the pure strategies case, the relationship, Maximin expected payoff ≤ Minimax expected payof.

Example: Pure Strategy in Game Theory

		Player B				
		I	**II**	**III**	**IV**	**V**
	I	-2	0	0	5	3
Player A	**II**	4	2	1	3	2
	III	-4	-3	0	-2	6
	IV	5	3	-4	2	-6

Solution.

We use the maximin (minimax) principle to analyze the game.

		Player B					
		I	**II**	**III**	**IV**	**V**	**Minimum**
	I	-2	0	0	5	3	-2
Player A	**II**	4	2	1	3	2	1
	III	-4	-3	0	-2	6	-4
	IV	5	3	-4	2	-6	-6
Maximum		5	3	1	5	6	

Select minimum from the maximum of columns.
Minimax = 1
Player A will choose II strategy, which yields the maximum payoff of 1.

Select maximum from the minimum of rows.

Maximin = 1
Similarly, player B will choose III strategy.

Since the value of maximin coincides with the value of the minimax, therefore, saddle point (equilibrium point) = 1.

The amount of payoff at an equilibrium point is also known as value of the game.

The optimal strategies for both players are: Player A must select II strategy and player B must select III strategy. The value of game is 1, which indicates that player A will gain 1 unit and player B will sacrifice 1 unit.

Mixed Strategy: Game Theory

Mixed strategy means a situation where a saddle point does not exist, the maximin (minimax) principle for solving a game problem breaks down. The concept is illustrated with the help of following example.

Example: Mixed Strategy in Game Theory

Two companies A and B are competing for the same product. Their different strategies are given in the following pay-off matrix:

Company B

		I	II	III
Company A	**I**	-2	14	-2
	II	-5	-6	-4
	III	-6	20	-8

Determine the optimal strategies for both the companies.

Solution.

First, we apply the maximin (minimax) principle to analyze the game.

		Company B			
		I	**II**	**III**	**Minimum**
Company A	**I**	-2	14	-2	-2
	II	-5	-6	-4	-6
	III	-6	20	-8	-8
Maximum		-2	20	-2	

Minimax = -2
Maximin = -2

There are two elements whose value is –2. Hence, the solution to such a game is not unique.

In the above problem, there is no saddle point. In such cases, the maximin and minimax principle of solving a game problem can't be applied. Under this situation, both the companies may resort to what is known as mixed strategy.

In a mixed strategy, each player moves in a random fashion.

DOMINANCE PROPERTY

The principle of dominance states that if one strategy of a player dominates over the other strategy in all conditions then the later strategy can be ignored. A strategy dominates over the other only if it is preferable over other in all conditions. The concept of dominance is especially useful for the evaluation of two-person zero-sum games where a saddle point does not exist.

In case of pay-off matrices larger than 2 × 2 size, the dominance property can be used to reduce the size of the pay-off matrix by eliminating the strategies that would never be selected.

Example : Solve the game given below in Table after reducing it to 2 × 2 game:

Game Problem

$$
\begin{array}{cc}
 & \textbf{Player B} \\
 & \begin{array}{ccc} \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\
\textbf{Player A} \begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} & \begin{pmatrix} 1 & 7 & 2 \\ 6 & 2 & 7 \\ 5 & 1 & 6 \end{pmatrix}
\end{array}
$$

Solution: Reduce the matrix by using the dominance property. In the given matrix for player A, all the elements in Row 3 are less than the adjacent elements of Row 2. Strategy 3 will not be selected by player A, because it gives less profit for player A. Row 3 is dominated by Row 2. Hence delete Row 3, as shown in table.

Reduced the Matrix by Using Dominance Property

$$
\begin{array}{cc}
 & \textbf{Player B} \\
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\textbf{Player A} \begin{array}{c} \mathbf{1} \\ \mathbf{2} \end{array} & \begin{pmatrix} 1 & 7 & 2 \\ 6 & 2 & 7 \end{pmatrix}
\end{array}
$$

For Player B, Column 3 is dominated by column 1 (Here the dominance is opposite because Player B selects the minimum loss). Hence delete Column 3. We get the reduced 2 × 2 matrix as shown below in table.

Reduced 2 × 2 Matrix

$$
\begin{array}{c c}
 & \textbf{Player B} \\
 & \begin{array}{cc} 1 & 2 \end{array} \\
\text{Player A} \begin{array}{c} \mathbf{1} \\ \mathbf{2} \end{array} & \begin{pmatrix} 1 & 7 \\ 6 & 2 \end{pmatrix}
\end{array}
$$

Now, solve the 2 × 2 matrix, using the maximin criteria as shown below in table.

Maximin Procedure

$$
\begin{array}{l c c c}
 & \multicolumn{2}{c}{\textbf{Player B}} & \\
 & \mathbf{1} & \mathbf{2} & \textbf{Row Min} \\
\text{Player A} \quad \mathbf{1} & 1 & 7 & 1 \\
\phantom{\text{Player A}} \quad \mathbf{2} & 6 & 2 & 2 \\
\textbf{Column Max} & 6 & 7 &
\end{array}
$$

$$\text{Max Min} \neq \text{Min Max}$$

$$2 \neq 6$$

Therefore, there is no saddle point and the game has a mixed strategy. Applying the probability formula,

p1 =2-6(1+2)-(7+6) =-43-13 =410 =25 q1 =2-7(1+2)-(7+6) =-53-13 =510 =12q2 =1- q1 =1-12 =12 Value of the game, v =(1*2)-(7*6)(1+2)-(7+6) =2-423-13 =4010 = 4

The optimum strategies are shown in table

Optimum Strategies

(a) $S_A = \begin{bmatrix} A_1 & A_2 & A_3 \\ {}^{2}/_{5} & {}^{3}/_{5} & 0 \end{bmatrix}$ and (b) $S_B = \begin{bmatrix} B_1 & B_2 & B_3 \\ ½ & ½ & 0 \end{bmatrix}$

Value of the game, v = 4

Example : Is the following two-person zero-sum game stable? Solve the game given below in table.

Two-person Zero-sum Game Problem

$$
\begin{array}{ll}
 & \textbf{Player B} \\
 & \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\
\textbf{Player A} \begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{array} & \begin{pmatrix} 5 & -10 & 9 & 0 \\ 6 & 7 & 8 & 1 \\ 8 & 7 & 15 & 1 \\ 3 & 4 & -1 & 4 \end{pmatrix}
\end{array}
$$

Solution: Solve the given matrix using the maximin criteria as shown in table.

Table 14.25: Maximin Procedure

		Player B				
		1	**2**	**3**	**4**	**Row Min**
	1	5	–10	9	0	–10
Player A	**2**	6	7	8	1	1
	3	8	7	15	1	1
	4	3	4	–1	4	–1
Column Max		8	7	15	4	

Therefore, there is no saddle point and hence it has a mixed strategy.

The pay-off matrix is reduced to 2×2 size using dominance property. Compare the rows to find the row which dominates other row. Here for Player A, Row 1 is dominated by Row 3 (or row 1 gives the minimum profit for Player A), hence delete Row 1. The matrix is reduced as shown in table.

Use Dominance Property to Reduce Matrix (Deleted Row 1)

$$
\begin{array}{ll}
 & \textbf{Player B} \\
 & \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\
\textbf{Player A} \begin{array}{c} \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{array} & \begin{pmatrix} 6 & 7 & 8 & 1 \\ 8 & 7 & 15 & 1 \\ 3 & 4 & -1 & 4 \end{pmatrix}
\end{array}
$$

When comparing column wise, column 2 is dominated by column 4. For Player B, the minimum profit column is column 2, hence delete column 2. The matrix is further reduced as shown in table.

Matrix Further Reduced to 3×3 (2 Deleted Column)

$$
\begin{array}{ccc}
 & & \textbf{Player B} \\
 & & \begin{array}{ccc} \mathbf{1} & \mathbf{3} & \mathbf{4} \end{array} \\
\textbf{Player A} & \begin{array}{c} \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{array} & \begin{pmatrix} 6 & 8 & 1 \\ 8 & 15 & 1 \\ 3 & -1 & 4 \end{pmatrix}
\end{array}
$$

Now, Row 2 is dominated by Row 3, hence delete Row 2, as shown in table.

Reduced Matrix (Row 2 Deleted)

$$
\begin{array}{ccc}
 & & \textbf{Player B} \\
 & & \begin{array}{ccc} \mathbf{1} & \mathbf{3} & \mathbf{4} \end{array} \\
\textbf{Player A} & \begin{array}{c} \mathbf{3} \\ \mathbf{4} \end{array} & \begin{pmatrix} 8 & 15 & 1 \\ 3 & -1 & 4 \end{pmatrix}
\end{array}
$$

Now, as when comparing rows and columns, no column or row dominates the other. Since there is a tie while comparing the rows or columns, take the average of any two rows and compare. We have the following three combinations of matrices as shown in table.

Matrix Combinations

$$
\begin{array}{cc}
 & \text{(a) B} \\
 & \begin{array}{cc} \frac{R_1+R_3}{2} & R_3 \end{array} \\
A & \begin{pmatrix} 11.5 & 1 \\ 1 & 4 \end{pmatrix} \\
 & \times
\end{array}
\qquad
\begin{array}{cc}
 & \text{(b) B} \\
 & \begin{array}{cc} R_1 & \frac{R_3+R_4}{2} \end{array} \\
A & \begin{pmatrix} 8 & 8 \\ 3 & 1.5 \end{pmatrix} \\
 & \checkmark
\end{array}
\qquad
\begin{array}{cc}
 & \text{(c) B} \\
 & \begin{array}{cc} R_2 & \frac{R_1+R_4}{2} \end{array} \\
A & \begin{pmatrix} 15 & 4.5 \\ -1 & 3.5 \end{pmatrix} \\
 & \times
\end{array}
$$

When comparing column 1 and the average of column 3 and column 4, column 1 is dominated by the average of column 3 and 4. Hence delete column 1. Finally, we get the 2 × 2 matrix as shown in table.

2×2 Matrix After Deleting Column 1

$$\begin{array}{lcc} & & \textbf{Player B} \\ & & \begin{array}{cc} \mathbf{3} & \mathbf{4} \end{array} \\ \textbf{Player A} & \begin{array}{c} \mathbf{3} \\ \mathbf{4} \end{array} & \begin{pmatrix} 15 & 1 \\ -1 & 4 \end{pmatrix} \end{array}$$

The strategy for the arrived matrix is a mixed strategy; using probability formula, we find p1, p2 and q1, q2.

p1 =4-(-1)(15+4)-(1+(-1)) =519 p2= 1- =59 =149 q14-119 =319 q2 = 1-319 =1619 Value of the game, v =(15*4)-(1*(-1))(15+4)-(1+(-1))=60+119 =6119

The optimum mixed strategies are given below in table.

Optimum Mixed Strategies

(a) $S_A = \begin{pmatrix} A_1 & A_2 & A_3 & A_4 \\ 0 & 0 & 5/19 & 14/19 \end{pmatrix}$ and (b) $S_B = \begin{pmatrix} B_1 & B_2 & B_3 & B_4 \\ 0 & 0 & 3/19 & 16/19 \end{pmatrix}$

We will now attempt to find an optimal solution to the linear programming model we introduced in the previous section. The method we will employ is known as the *graphical method*and can be applied to any problem with two decision variables. It basically consists of two steps: Finding the *feasible region* or the *feasible space* (which is the region in the plane where all the feasible solutions to the problems lie) and then identifying the optimal solution among all the feasible ones.

To begin the procedure we first graph the lines

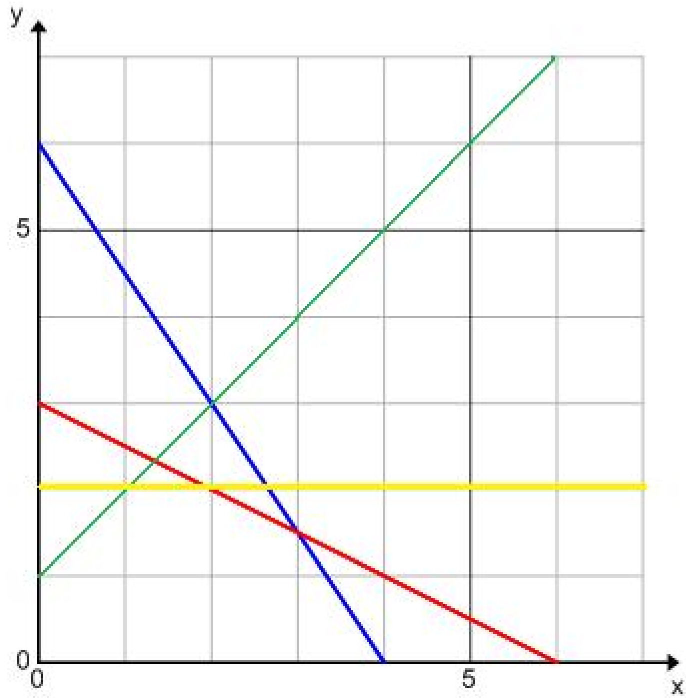
,

,

,

,

in the first quadrant. Note that for our purpose and on the graph.



We now will shade the feasible region. To do that consider the constraints one by one. The first one

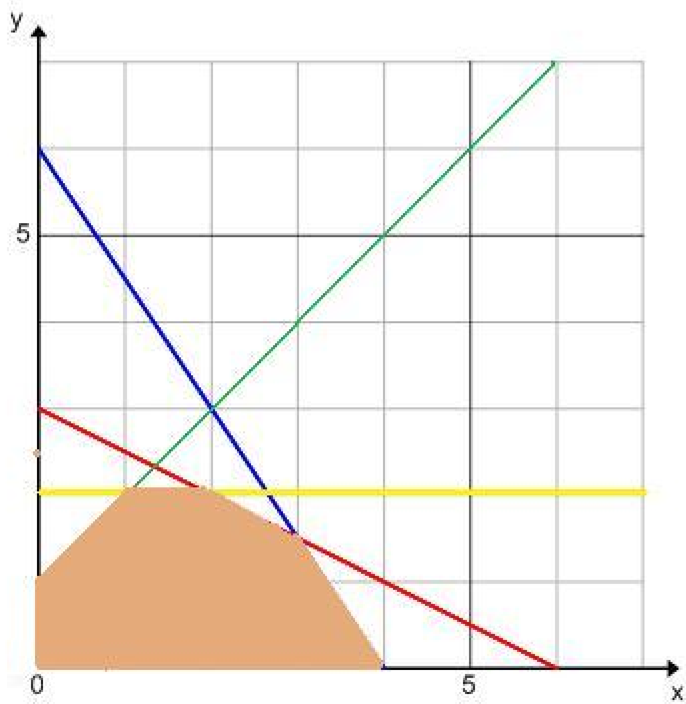
is . To determine the region it represents, choose any point which does not pass through the

line say (0,0). Substitute it in the constraint to get . Since this is true we conclude

that (0,0) lies in the region represented by . We conclude that all the points on the side of

containing (0,0) actually represent . This is suggested by the fact that the line divides the plane in 2 distinct halfs: One of points satisfying the inequality and one of those which don't.

In this way all the inequalities can be shaded. The region which is shaded under all inequalities is the feasible region of the whole problem. Clearly in this region all the constraints of the problem hold. (The non negativity restrictions hold since we are working in the first quadrant.)



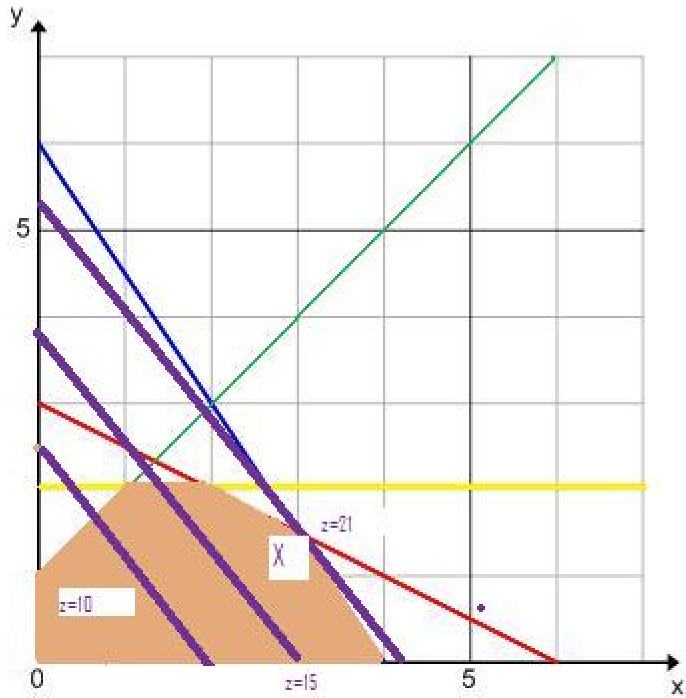
We are now ready to find out the optimal solutions. To do this graph the line . Since

represents the objective function so represents the points where the objective function has

value 10 (i.e. the total profit is 10). Now plot the line which represents the points where the objective function has value 15. This gives us an idea of the direction of increase in z. The optimal solution occurs at the point X which is the point beyond which any further increase will put z outside

the boundaries of the feasible region. The coordiantes of X can be found by solving and

so that and . This is the optimal solution to the problem and indicates that the amounts of salts X and Y should be 3 and 1.5 respectively. This will give the maximum profit of 21 which is the *optimal value*.



A point to note is that the optimal solution in a LP model always occurs at a *corner point* of a feasible region. This is true even if the line z=c comes out to be parallel to one of the constraints. Although a mathematical proof of this fact would involve considerable linear algebra we will satisfy ourselves of it by noting that that any objective function in the feasible region would have glided out of the region just after touching one of the corner points.

A minimization example

Let us look at a minimization problem. It can occur in actual practice when instead of the profits associated with the salts X and Y we are given the costs of their production. All we have to do is now move the line z=c in the direction of its decrease and we have the optimal solution at the point ((0,0) in our example) where any further decrease will take z outside the feasible region. Another way to solve the problem is to convert the min problem into a max one. To do that simple consider the negative of the objective function.

SM

We can solve two variables LP models easily using the graphical method outlined in the previous section but what should we do in case of three variable problems, i.e. when our company makes three products we have to make decisions about. What about four or five variable problems? This is where the simplex method comes in. It is an iterative method which by repeated use gives us the solution to any n variable LP model.

Let us define an LP model as being in standard form if it satisfies the following two conditions:

- All the constraints with the exception of the non-negativity restrictions are equations with non-negative right hand side.

- All the variables are non-negative.

For example,

Maximize
subject to,

,

,

,

,

is an LP model in standard form. Note that the usual signs have been equivalently

replaced by both variables (s as in **s**lack) and the sign.

How do we convert a given model into a standard model while retaining its sense? This is done as follows:

- If there is a constraint of the type then the right hand side usually represents some limit on the resource (for example the amount of the raw material available) and the left hand side the usage of that resource (i.e. how much raw material is actually used) and so their difference represents the *slack* or unused amount of the resource. So to convert into an equality we must add a variable representing the slack amount on the left hand side. This variable is known as the *slack variable* and is obviously also non-negative. If

  we represent it by our constraint is converted into the

  equation . A similar thing is done in case of a constraint of the

  type . Here the left hand side has a *surplus* or extra amount then the right hand side and so a non-negative *surplus*

  variable say must be subtracted to get the equation .

- If the given variable is given to be non-positive then its

  negative is obviously non-negative and can be substituted in the problem. The real problem comes in the case when the variable is allowed to take on any sign. Then it is called an *unrestricted*

  variable. This is overcome by using the substitution where are both non-negative. Intuitively if the variable is

positive then is positive and is zero, while if the

variable is negative then is zero and is positive. If

is zero then obviously are both zero.

A point to note is that the objective function in the original LP model and the standard model is the same.

Algebraic solution of a LP

Let us now try to analyze a LP model algebraically. A solution of the standard LP model will be a solution for the original model (since the slack and surplus variables once removed would make the equations regain their old imbalance) and due to a similar reasoning a solution for the original model will also correspond to a solution for the standard model. Since the objective function is the same for both models so an optimal solution for the standard model will be optimal for the original model as well. Therefore we need to bother only about the standard model.

Now what are the candidates for the optimal solution? They are the solutions of the equality constraints. All we need to do is to find out the solutions and check which of those give the optimal value when put in the objective function. Now usually in the LP model the number of constraints, say m, is outnumbered by the number of variables, say n, and so there are infinitely many solutions to the system of constraints. We can't possibly examine the complete set of infinite solutions. However due to a mathematical result our work is shortened. The result is that if out of the n variables, n-m variables are put to zero, and then if the constraint system can be solved then the solution will correspond to a corner point in the n-space. Such a solution is called a *basic solution (Initial Solution)*. If in addition to being basic it happens to be feasible to the original problem, then it is called a *basic feasible solution* (often abbreviated as BFS). Since the optimal solution is obtained on a corner point (as we observed graphically) so all we need to do is to examine all the basic feasible

solutions (which are at most in number, reflective of the number of ways n-m variables can be chosen among the total n to be zero) and then decide which one gives the maximum value for the objective function.

Let us consider an example:

Consider the following LP model:

Maximize z =

subject to,

,

,

We first convert it into standard form:

Maximize z =

subject to,

The constraint system has m=2 constraints and n=4 variables. Thus we need to set 4-2=2 variables equal to zero to get a basic solution. For

example putting and we get a solution and . This is also clearly feasible and so is a basic feasible solution. The variables

being put to zero, that are are called *nonbasicvariables*and are called *basic variables*. The *objective value* (the value that is obtained by putting the solution in the objective function) that these solutions (both basic and non basic) give on substitution in the objective function is zero.

The following table summarizes all the basic solutions to the problem:

Nonbasic variables	**Basic variables**	**Basic solution**	**Feasible**	**Objective value**
		(4,5)	Yes	0
		(4,-3)	No	-
		(2.5,1.5)	Yes	7.5
		(2,3)	Yes	4
		(5,-6)	No	-
		(1,2)	Yes	8

Note that we have not bothered to calculate the objective value for infeasible solutions. The optimal solution is the one which yields the

highest objective value i.e. . Hence we have solved the LP model algebraically.

This procedure of solving LP models works for any number of variables but is very difficult to employ when there are a large number of constraints and variables. For example, for m = 10 and n = 20 it is

necessary to solve sets of equations, which is clearly a staggering task. With the simplex method, you need only solve a few of these sets of equations, concentrating on those which give improving objective values.

The Simplex method

The method in a nutshell is this. You start with a basic feasible solution of an LP in standard form (usually the one where all the slack variables are equal to the corresponding right hand sides and all other variables are zero) and replace one basic variable with one which is currently non-basic to get a new basic solution (since n-m variables remain zero). This is done in a fashion which ensures that the new basic solution is feasible and its objective value is at least as much as that of the previous BFS. This is repeated until it is clear that the current BFS can't be improved for a better objective value. In this way the optimal solution is achieved.

It is clear that one factor is crucial to the method: which variable should replace which. The variable which is replaced is called the *leaving variable* and the variable which replaces it is known as the *entering variable*. The design of the simplex method is such so that the process of choosing these two variables allows two things to happen. Firstly, the new objective value is an improvement(or at least equals) on the current one and secondly the new solution is feasible.

Let us now explain the method through an example. Consider our old chemical company problem in standard form:

Maximize z =

subject to,

Now an immediate BFS is obtained by putting all the equal to zero. (Clearly the solution thus obtained will be feasible to the original problem as the right hand sides are all non-negative which is precisely

our solution.) If we consider our objective function z = then it is

evident that an increase in or will increase our objective value. (Note that currently both are zero being non-basic). A unit

increase in will give a 5-fold increase in the objective value while a unit increase in will give a 4-fold increase. It is logical that we

elect to make the entering variable in the next iteration.

In the tabular form of the simplex method the objective function is

usually represented as . Also the table contains the system of constraints along with the BFS that is obtained. Only the coefficients are written as is usual when handling linear systems.

Basic	**z**							**BFS**
Z	1	-5	-4	0	0	0	0	0
	0	6	4	1	0	0	0	24
	0	1	2	0	1	0	0	6
	0	-1	1	0	0	1	0	1
	0	0	1	0	0	0	1	2

Now for the next iteration we have to decide the entering and the

leaving variables. The entering variable is as we discussed. In fact, due to our realignment of the objective function, the most negative value in the z-row of the simplex table will always be the entering variable for the next iteration. This is known as the *optimality condition*. What about the leaving variable? We have to account for the fact that our next basic solution is feasible. So our leaving variable must be chosen with this thought in mind.

To decide the leaving variable we apply what is sometimes called as the *feasibility condition*. That is as follows: we compute the quotient of the solution coordinates (that are 24, 6, 1 and 2) with the constraint

coefficients of the entering variable (that are 6, 1, -1 and 0). The following ratios are obtained: 24/6 = 4, 6/1 = 6, 1/-1 = -1 and 2/0 = undefined. Ignoring the negative and the undefined ratio we now proceed to select the minimum of the other two ratios which is 4,

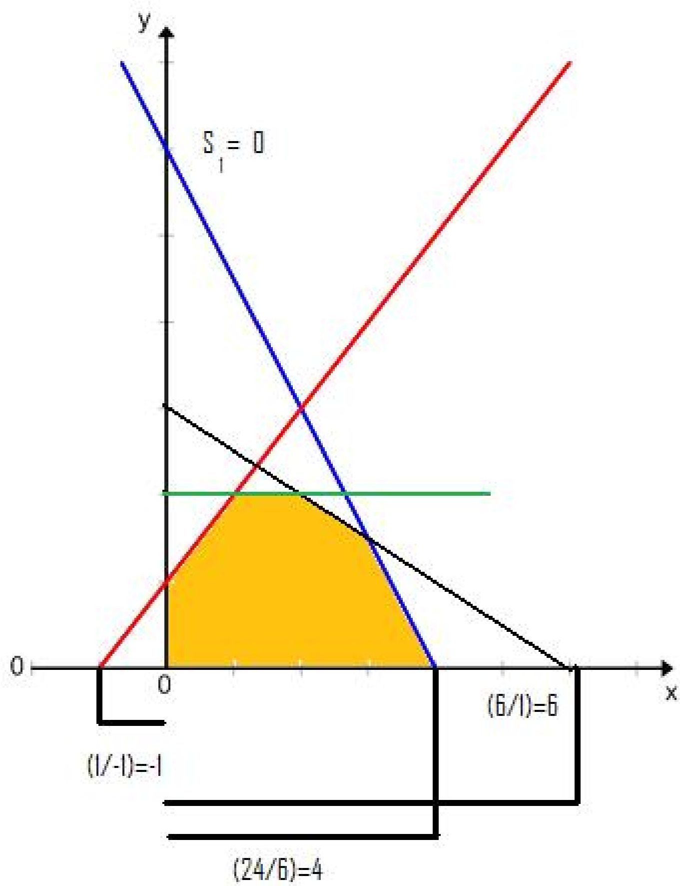
obtained from dividing 24(the value of) by 6. Since the minimum

involved the division by 's current value we take the leaving

variable as .

What is the justification behind this procedure? It is this. The minimum ratios actually represent the intercepts made by the constraints on the

axis. To see this look at the following graph:



Since currently all are 0 we are considering the BFS corresponding to the origin. Now, in the next iteration according to the simplex method we should get a new BFS i.e move to a new corner point on the graph. We can induce an increase in the value of only one

variable at a time by making it an entering variable, and since is our entering variable our plan is to increase the value of . From the

graph we can see that the value of must be increased to 4 at the

point (4,0), which is the smallest non-negative intercept with the axis. An increase beyond that point is infeasible. Also at (4,0) the slack

variable assumes a zero value as the first constraint is satisfied

as an equality there and so becomes the leaving variable.

Now the problem is to determine the new solution. Although any procedure of solving a linear system can be applied at this stage, usually *Gauss Jordan elimination* is applied. It is based on a result in linear algebra that the elementary row transformations on a system [A|b] to [H|c] do not alter the solutions of the system. According to it the columns corresponding to the basic-variables in the table are given the shape of an identity matrix. (Readers familiar with linear algebra will recognize that it means that the matrix formed with the basis variable columns is transformed into reduced row echelon form.) The solution can then be simply read off from the right most solution column (as n-m of the variables are put to zero and the rest including z have coefficient 1 in one constraint each). Since z is also a variable it's row is treated as one among the constraints comprising the linear system.

The entering variable column is called the *pivot column* and the leaving variable row is called the *pivot row*. The intersection of the pivot column and the leaving variable column is called the *pivot element*. In our

example the second row (of) is the pivot row and the second

column(of) is the pivot column.

The computations needed to produce the new basic solution are:

- Replace the leaving variable in the 'Basic' column with the entering variable.
- New pivot row = Current pivot row ÷ Pivot element

For other rows, including the z row:

- New row = Current row - (Its pivot column coefficient)×(New pivot row)

In our case the computations proceed as follows:

- Replace in the 'Basic' column with
- New pivot row = Current row ÷ 6 =
- New z row = Current z row - (-5)×New row =

- New row = Current row - (1)×New row =
- New row = Current row - (-1)×New row =
- New row = Current row - (0)×New row =

Thus our new table is:

Basic	z							BFS
Z	1	0			0	0	0	20
	0	1			0	0	0	4
	0	0			1	0	0	2
	0	0			0	1	0	5
	0	0	1	0	0	0	1	2

The optimality condition now shows that is the entering variable. The feasiblity condition produces the ratios: 4/(2/3) = 6, 2/(4/3) = 1.5, 5/(5/3) = 3 and 2/1 = 2 of which the minimum is 1.5 produced by

dividing the coefficient in the row (i.e. the row in which the basic

variable has coefficient 1). So becomes the leaving variable.

The Gauss Jordan elimination process is applied again to get the following table:

Basic	z							BFS
Z	1	0	0			0	0	21
	0	1	0			0	0	3
	0	0	1			0	0	
	0	0	0			1	0	
	0	0	0			0	1	

Based on the optimality condition none of the z-row coefficients

associated with the non-basic variables and are negative. Hence the last table is optimal. The optimal solution can thus be read

off as and and the optimal value as z = 21. Obviously the

variables and are zero. Our original problem involving only

and also clearly has the same solution (just disregard the slacks).

We have dealt with a maximization problem. If the minimization case, since min z = max (-z) (if z is a linear function, which it is) so we can either convert the problem into a maximization one or reverse the optimality condition. Instead of selecting the entering variable as the one with the most negative coefficient in the z row we can select the one with the most positive coefficient. The rest of the steps are the same.

EXERCISES

1. What is Decision-making? Explain and differentiate this under the conditions of certainty and uncertainty.

2. What are different environments in which decisions are made?

3. What is EMV? How is it computed to be used as a criterion of decision-making and when?

4. Write short note on the value of perfect information.

5. What do you mean by decision tree analysis? What is node in a decision tree? What is backward pass?

6. What is the concept of decision tree analysis? What are the basic steps involved in the construction of such tree?

7. Explain the construction of decision tree. What do you mean by it?

8. A businessman has two independent investments X and Y available to him, but he lacks the capital to undertake both of them simultaneously. He can choose to take X first and then stop, or if X is successful, then take Y or vice versa. The probability of success of X is 0.7, while for Y it is 0.4. Both investments require an initial capital outlay of Rs.2,000and both return nothing if the venture is unsuccessful. Successful completion of X will return Rs.3,000(over cost) and successful completion of Y will return Rs.5,000 (over cost). Draw the decision tree and dertermine the best strategy.

9. A manager has a choice between (a) A risky contract promising Rs.7 lakhs with probability 0.6 and Rs. 4 lakhs with probability 0.4. (b) A diversified portfolio consisting of two contracts with independent outcomes and each promising Rs. 3.5 lakhs with probability 0.6 and Rs. 2 lakhs with probability 0.4. Construct a decision tree using EMV criterion. Can you arrive at the decision using EMV criterion?

10. Write short note on applications of game theory.

11. What do you understand by zero-sum and nonzero-sum games?

12. What do you mean by strategy, dominance and saddle point?

13. Explain the following:

    a) Minimax and maximin principles

    b) Pure and mixed strategies

    c) Teo-person zeros-sum game.

14. Discuss various methods of finding solutions to a given game.

15. Describe the graphical method to solve games.

16. Discuss the algebraic method for solving 2 x 2 games by taking a suitable example.

17. Show how a game can be formulated as an L.P.P.

18. Write any four limitations of game theory.

19. Discuss the role of theory of games for scientific decision-making

20. Explain the assumptions underlying game theory.

21. State and prove minimax theorem for two-person zero-sum games.

UNIT – V

Queuing Theory

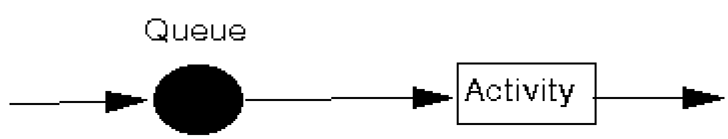
Queuing theory deals with problems which involve queuing (or waiting). Typical examples might be:

- banks/supermarkets - waiting for service
- computers - waiting for a response
- failure situations - waiting for a failure to occur e.g. in a piece of machinery
- public transport - waiting for a train or a bus

As we know queues are a common every-day experience. Queues form because resources are limited. In fact it makes *economic sense* to have queues. For example how many supermarket tills you would need to avoid queuing? How many buses or trains would be needed if queues were to be avoided/eliminated?

In designing queueing systems we need to aim for a balance between service to customers (short queues implying many servers) and economic considerations (not too many servers).

In essence all queuing systems can be broken down into individual sub-systems consisting of *entities* queuing for some *activity* (as shown below).



Typically we can talk of this individual sub-system as dealing with ***customers*** queuing for ***service***. To analyse this sub-system we need information relating to:

- **arrival process:**
  - how customers arrive e.g. singly or in groups (batch or bulk arrivals)
  - how the arrivals are distributed in time (e.g. what is the probability distribution of time between successive arrivals (the ***interarrival time distribution***))
  - whether there is a finite population of customers or (effectively) an infinite number

The simplest arrival process is one where we have completely regular arrivals (i.e. the same constant time interval between successive arrivals). A Poisson stream of arrivals corresponds to arrivals at random. In a Poisson stream successive customers arrive after intervals which independently are exponentially distributed. The Poisson stream is important as it is a convenient mathematical model of many real life queuing systems and is described by a single parameter - the average arrival rate. Other important arrival processes are scheduled arrivals; batch arrivals; and time dependent arrival rates (i.e. the arrival rate varies according to the time of day).

- **service mechanism:**
    - a description of the resources needed for service to begin
    - how long the service will take (the ***service time distribution***)
    - the number of servers available
    - whether the servers are in series (each server has a separate queue) or in parallel (one queue for all servers)
    - whether preemption is allowed (a server can stop processing a customer to deal with another "emergency" customer)

    Assuming that the service times for customers are independent and do not depend upon the arrival process is common. Another common assumption about service times is that they are exponentially distributed.

- **queue characteristics:**
    - how, from the set of customers waiting for service, do we choose the one to be served next (e.g. FIFO (first-in first-out) - also known as FCFS (first-come first served); LIFO (last-in first-out); randomly) (this is often called the *queue discipline*)
    - do we have:
        - balking (customers deciding not to join the queue if it is too long)
        - reneging (customers leave the queue if they have waited too long for service)
        - jockeying (customers switch between queues if they think they will get served faster by so doing)
        - a queue of finite capacity or (effectively) of infinite capacity

> Changing the queue discipline (the rule by which we select the next customer to be served) can often reduce congestion. Often the queue discipline "choose the customer with the lowest service time" results in the smallest value for the time (on average) a customer spends queuing.

Note here that integral to queuing situations is the idea of uncertainty in, for example, interarrival times and service times. This means that probability and statistics are needed to analyse queuing situations.

In terms of the analysis of queuing situations the types of questions in which we are interested are typically concerned with measures of system performance and might include:

- How long does a customer expect to wait in the queue before they are served, and how long will they have to wait before the service is complete?
- What is the probability of a customer having to wait longer than a given time interval before they are served?
- What is the average length of the queue?
- What is the probability that the queue will exceed a certain length?
- What is the expected utilisation of the server and the expected time period during which he will be fully occupied (remember servers cost us money so we need to keep them busy). In fact if we can assign costs to factors such as customer waiting time and server idle time then we can investigate how to design a system at minimum total cost.

These are questions that need to be answered so that management can evaluate alternatives in an attempt to control/improve the situation. Some of the problems that are often investigated in practice are:

- Is it worthwhile to invest effort in reducing the service time?
- How many servers should be employed?
- Should priorities for certain types of customers be introduced?
- Is the waiting area for customers adequate?

In order to get answers to the above questions there are *two* basic approaches:

- analytic methods or queuing theory (formula based); and
- simulation (computer based).

The reason for there being two approaches (instead of just one) is that analytic methods are only available for relatively simple queuing systems. Complex queuing systems are almost always analysed using simulation (more technically known as discrete-event simulation).

The simple queueing systems that can be tackled via queueing theory essentially:

- consist of just a single queue; linked systems where customers pass from one queue to another cannot be tackled via queueing theory
- have distributions for the arrival and service processes that are well defined (e.g. standard statistical distributions such as Poisson or Normal); systems where these distributions are derived from observed data, or are time dependent, are difficult to analyse via queueing theory

The first queueing theory problem was considered by Erlang in 1908 who looked at how large a telephone exchange needed to be in order to keep to a reasonable value the number of telephone calls not connected because the exchange was busy (lost calls). Within ten years he had developed a (complex) formula to solve the problem.

Additional queueing theory information can be found here and here

---

Queueing notation and a simple example

It is common to use to use the symbols:

- *lamda* to be the mean (or average) number of arrivals per time period, i.e. the mean arrival rate
- µ to be the mean (or average) number of customers served per time period, i.e. the mean service rate

There is a standard notation system to classify queueing systems as A/B/C/D/E, where:

- A represents the probability distribution for the arrival process
- B represents the probability distribution for the service process
- C represents the number of channels (servers)
- D represents the maximum number of customers allowed in the queueing system (either being served or waiting for service)
- E represents the maximum number of customers in total

Common options for A and B are:

- M for a Poisson arrival distribution (exponential interarrival distribution) or a exponential service time distribution
- D for a deterministic or constant value
- G for a general distribution (but with a known mean and variance)

If D and E are not specified then it is assumed that they are infinite.

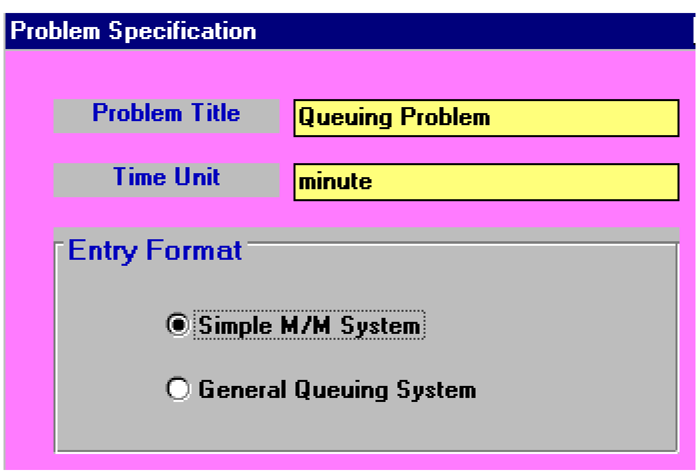
For example the M/M/1 queueing system, the simplest queueing system, has a Poisson arrival distribution, an exponential service time distribution and a single channel (one server).

Note here that in using this notation it is always assumed that there is just a **single queue** (waiting line) and customers move from this single queue to the servers.

Simple M/M/1 example

Suppose we have a single server in a shop and customers arrive in the shop with a Poisson arrival distribution at a mean rate of *lamda*=0.5 customers per minute, i.e. on average one customer appears every 1/*lamda* = 1/0.5 = 2 minutes. This implies that the interarrival times have an exponential distribution with an average interarrival time of 2 minutes. The server has an exponential service time distribution with a mean service rate of 4 customers per minute, i.e. the service rate μ=4 customers per minute. As we have a Poisson arrival rate/exponential service time/single server we have a M/M/1 queue in terms of the standard notation.

We can analyse this queueing situation using the package. The input is shown below:



Data Description	ENTRY
Number of servers	1
Service rate (per server per minute)	4
Customer arrival rate (per minute)	0.5
Queue capacity (maximum waiting space)	M
Customer population	M

with the output being:

11-14-2000	Performance Measure	Result
1	System: M/M/1	From Formula
2	Customer arrival rate (lambda) per minute =	0.5000
3	Service rate per server (mu) per minute =	4.0000
4	Overall system effective arrival rate per minute =	0.5000
5	Overall system effective service rate per minute =	0.5000
6	Overall system utilization =	12.5000 %
7	Average number of customers in the system (L) =	0.1429
8	Average number of customers in the queue (Lq) =	0.0179
9	Average number of customers in the queue for a busy system (Lb) =	0.1429
10	Average time customer spends in the system (W) =	0.2857 minutes
11	Average time customer spends in the queue (Wq) =	0.0357 minutes
12	Average time customer spends in the queue for a busy system (Wb) =	0.2857 minutes
13	The probability that all servers are idle (Po) =	87.5000 %
14	The probability an arriving customer waits (Pw or Pb) =	12.5000 %

The first line of the output says that the results are from a formula. For this very simple queueing system there are exact formulae that give the statistics above under the assumption that the system has reached a **steady state - that is that the system has been running long enough so as to settle down into some kind of equilibrium position.**

Naturally real-life systems hardly ever reach a steady state. Simply put life is not like that. However despite this simple queueing formulae can give us some insight into how a system might behave very quickly. The package took a fraction of a second to produce the output seen above.

One factor that is of note is *traffic intensity* = (arrival rate)/(departure rate) where arrival rate = number of arrivals per unit time and departure rate = number of departures per unit time. Traffic intensity is a measure of the congestion of the system. If it is near to zero there is very little queuing and in general as the traffic intensity increases (to near 1 or even greater than 1) the amount of queuing increases. For the system we have considered above the arrival rate is 0.5 and the departure rate is 4 so the traffic intensity is 0.5/4 = 0.125

Faster servers or more servers?

Consider the situation we had above - which would you prefer:

- one server working twice as fast; or
- two servers each working at the original rate?

The simple answer is that we can analyse this using the package. For the first situation one server working twice as fast corresponds to a service rate μ=8 customers per minute. The output for this situation is shown below.

11-15-2000	Performance Measure	Result
1	System: M/M/1	From Formula
2	Customer arrival rate (lambda) per minute =	0.5000
3	Service rate per server (mu) per minute =	8.0000
4	Overall system effective arrival rate per minute =	0.5000
5	Overall system effective service rate per minute =	0.5000
6	Overall system utilization =	6.2500 %
7	Average number of customers in the system (L) =	0.0667
8	Average number of customers in the queue (Lq) =	0.0042
9	Average number of customers in the queue for a busy system (Lb) =	0.0667
10	Average time customer spends in the system (W) =	0.1333 minutes
11	Average time customer spends in the queue (Wq) =	0.0083 minutes
12	Average time customer spends in the queue for a busy system (Wb) =	0.1333 minutes
13	The probability that all servers are idle (Po) =	93.7500 %
14	The probability an arriving customer waits (Pw or Pb) =	6.2500 %

For two servers working at the original rate the output is as below. Note here that this situation is a M/M/2 queueing system. Note too that the package assumes that these two servers are fed from a single queue (rather than each having their own individual queue).

Data Description	ENTRY
Number of servers	2
Service rate (per server per minute)	4
Customer arrival rate (per minute)	0.5
Queue capacity (maximum waiting space)	M
Customer population	M

11-15-2000	Performance Measure	Result
1	System: M/M/2	From Formula
2	Customer arrival rate (lambda) per minute =	0.5000
3	Service rate per server (mu) per minute =	4.0000
4	Overall system effective arrival rate per minute =	0.5000
5	Overall system effective service rate per minute =	0.5000
6	Overall system utilization =	6.2500 %
7	Average number of customers in the system (L) =	0.1255
8	Average number of customers in the queue (Lq) =	0.0005
9	Average number of customers in the queue for a busy system (Lb) =	0.0667
10	Average time customer spends in the system (W) =	0.2510 minutes
11	Average time customer spends in the queue (Wq) =	0.0010 minutes
12	Average time customer spends in the queue for a busy system (Wb) =	0.1333 minutes
13	The probability that all servers are idle (Po) =	88.2353 %
14	The probability an arriving customer waits (Pw or Pb) =	0.7353 %

Compare the two outputs above - which option do you prefer?

Of the figures in the outputs above some are identical. Extracting key figures which are different we have:

	One server twice as fast	Two servers, original rate
Average time in the system (waiting and being served)	0.1333	0.2510
Average time in the queue	0.0083	0.0010
Probability of having to wait for service	6.25%	0.7353%

It can be seen that with one server working twice as fast customers spend less time in the system on average, but have to wait longer for service and also have a higher probability of having to wait for service.

Extending the example: M/M/1 and M/M/2 with costs

Below we have extended the example we had before where now we have multiplied the customer arrival rate by a factor of six (i.e. customers arrive 6 times as fast as before). We have also entered a queue capacity (waiting space) of 2 - i.e. if all servers are occupied and 2 customers are waiting when a new customer appears then they go away - this is known as **balking**.

We have also added cost information relating to the server and customers:

- each minute a server is idle costs us £0.5
- each minute a customer waits for a server costs us £1
- each customer who is balked (goes away without being served) costs us £5

The package input is shown below:

Data Description	ENTRY
Number of servers	1
Service rate (per server per minute)	4
Customer arrival rate (per minute)	3
Queue capacity (maximum waiting space)	2
Customer population	M
Busy server cost per minute	
Idle server cost per minute	0.5
Customer waiting cost per minute	1
Customer being served cost per minute	
Cost of customer being balked	5
Unit queue capacity cost	

with the output being:

11-14-2000	Performance Measure	Result
1	System: M/M/1/3	From Formula
2	Customer arrival rate (lambda) per minute =	3.0000
3	Service rate per server (mu) per minute =	4.0000
4	Overall system effective arrival rate per minute =	2.5371
5	Overall system effective service rate per minute =	2.5371
6	Overall system utilization =	63.4286 %
7	Average number of customers in the system (L) =	1.1486
8	Average number of customers in the queue (Lq) =	0.5143
9	Average number of customers in the queue for a busy system (Lb) =	0.8108
10	Average time customer spends in the system (W) =	0.4527 minutes
11	Average time customer spends in the queue (Wq) =	0.2027 minutes
12	Average time customer spends in the queue for a busy system (Wb) =	0.3196 minutes
13	The probability that all servers are idle (Po) =	36.5714 %
14	The probability an arriving customer waits (Pw or Pb) =	63.4286 %
15	Average number of customers being balked per minute =	0.4629
16	Total cost of busy server per minute =	$0
17	Total cost of idle server per minute =	$0.1829
18	Total cost of customer waiting per minute =	$0.5143
19	Total cost of customer being served per minute =	$0
20	Total cost of customer being balked per minute =	$2.3143
21	Total queue space cost per minute =	$0
22	Total system cost per minute =	$3.0114

Note, as the above output indicates, that this is an M/M/1/3 system since we have 1 server and the maximum number of customers that can be in the system (either being served or waiting) is 3 (one being served, two waiting).

The key here is that as we have entered cost data we have a figure for the total cost of operating this system, 3.0114 per minute (in the steady state).

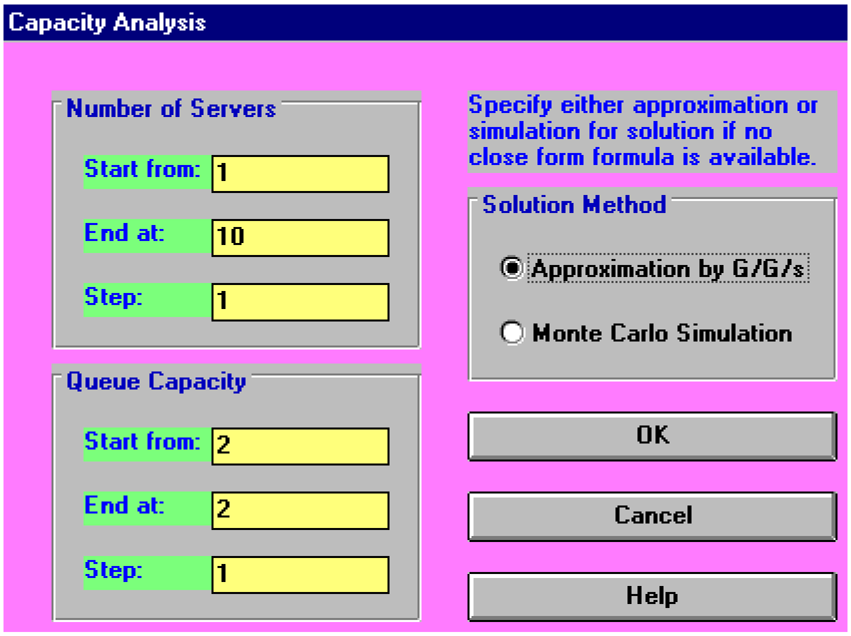
Suppose now we were to have two servers instead of one - would the cost be less or more? The simple answer is that the package can tell us, as below. Note that this is an M/M/2/4 queueing system as we have two servers and a total number of customers in the system of 4 (2 being served, 2 waiting in the queue for service). Note too that the package assumes that these two servers are fed from a single queue (rather than each having their own individual queue).

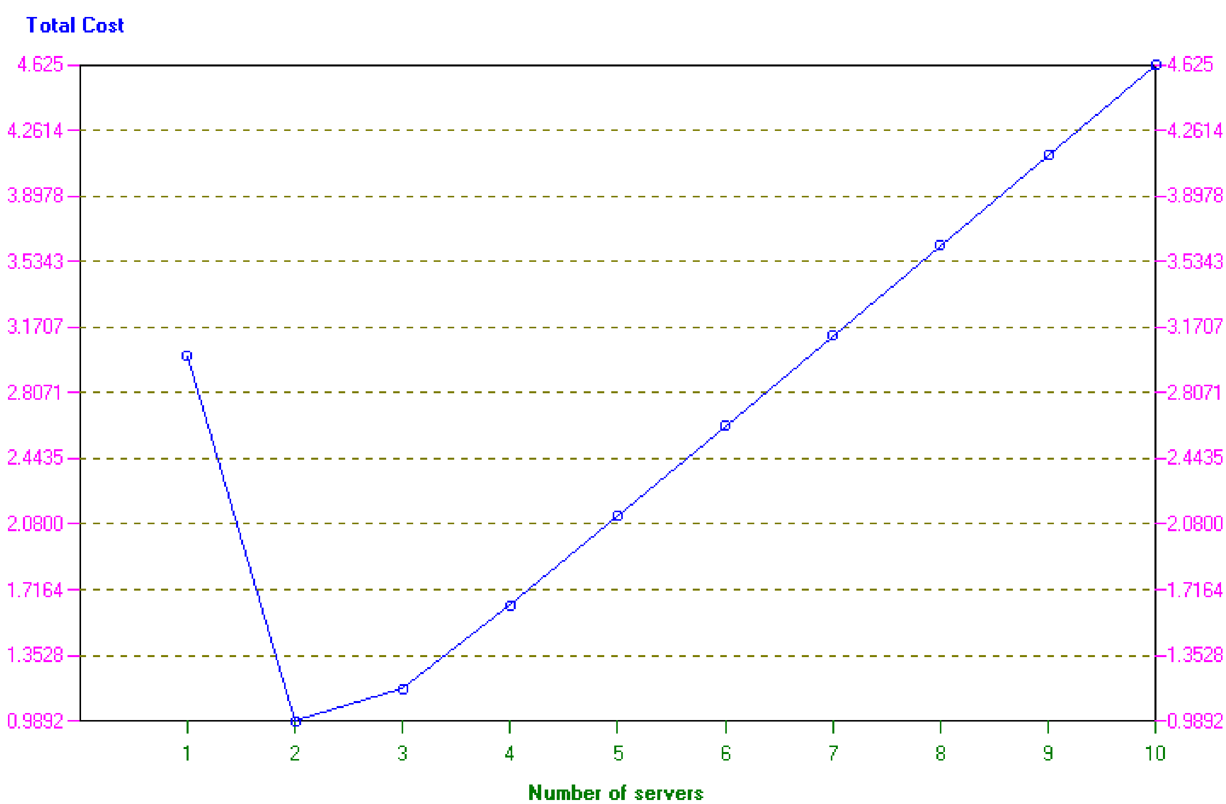
Data Description	ENTRY
Number of servers	2
Service rate (per server per minute)	4
Customer arrival rate (per minute)	3
Queue capacity (maximum waiting space)	2
Customer population	M
Busy server cost per minute	
Idle server cost per minute	0.5
Customer waiting cost per minute	1
Customer being served cost per minute	
Cost of customer being balked	5
Unit queue capacity cost	

11-14-2000	Performance Measure	Result
1	System: M/M/2/4	From Formula
2	Customer arrival rate (lambda) per minute =	3.0000
3	Service rate per server (mu) per minute =	4.0000
4	Overall system effective arrival rate per minute =	2.9455
5	Overall system effective service rate per minute =	2.9455
6	Overall system utilization =	36.8185 %
7	Average number of customers in the system (L) =	0.8212
8	Average number of customers in the queue (Lq) =	0.0848
9	Average number of customers in the queue for a busy system (Lb) =	0.4330
10	Average time customer spends in the system (W) =	0.2788 minutes
11	Average time customer spends in the queue (Wq) =	0.0288 minutes
12	Average time customer spends in the queue for a busy system (Wb) =	0.1470 minutes
13	The probability that all servers are idle (Po) =	45.9502 %
14	The probability an arriving customer waits (Pw or Pb) =	19.5872 %
15	Average number of customers being balked per minute =	0.0545
16	Total cost of busy server per minute =	$0
17	Total cost of idle server per minute =	$0.6318
18	Total cost of customer waiting per minute =	$0.0848
19	Total cost of customer being served per minute =	$0
20	Total cost of customer being balked per minute =	$0.2726
21	Total queue space cost per minute =	$0
22	Total system cost per minute =	$0.9892

So we can see that there is a considerable cost saving per minute in having two servers instead of one.

In fact the package can automatically perform an analysis for us of how total cost varies with the number of servers. This can be seen below.





General queueing

The screen below shows the possible input parameters to the package in the case of a general queueing model (i.e. not a M/M/r system).

Data Description	ENTRY
Number of servers	
Service time distribution (in minute)	Exponential
Location parameter (a)	
Scale parameter (b>0) (b=mean if a=0)	
(Not used)	
Service pressure coefficient	
Interarrival time distribution (in minute)	Exponential
Location parameter (a)	
Scale parameter (b>0) (b=mean if a=0)	
(Not used)	
Arrival discourage coefficient	
Batch (bulk) size distribution	Constant
Constant value	1
(Not used)	
(Not used)	
Queue capacity (maximum waiting space)	M
Customer population	M
Busy server cost per minute	
Idle server cost per minute	
Customer waiting cost per minute	
Customer being served cost per minute	
Cost of customer being balked	
Unit queue capacity cost	

Here we have a number of possible choices for the service time distribution and the interarrival time distribution. In fact the package recognises some 15 different distributions! Other items mentioned above are:

- service pressure coefficient - indicates how servers speed up service when the system is busy, i.e. when all servers are busy the service rate is increased. If this coefficient is s and we have r servers each with service rate µ then the service rate changes from µ to $(n/r)^s\mu$ when there are n customers in the system and n>=r.
- arrival discourage coefficient - indicates how customer arrivals are discouraged when the system is busy, i.e. when all servers are busy the arrival rate is decreased. If this coefficient is s and we have r servers with the arrival

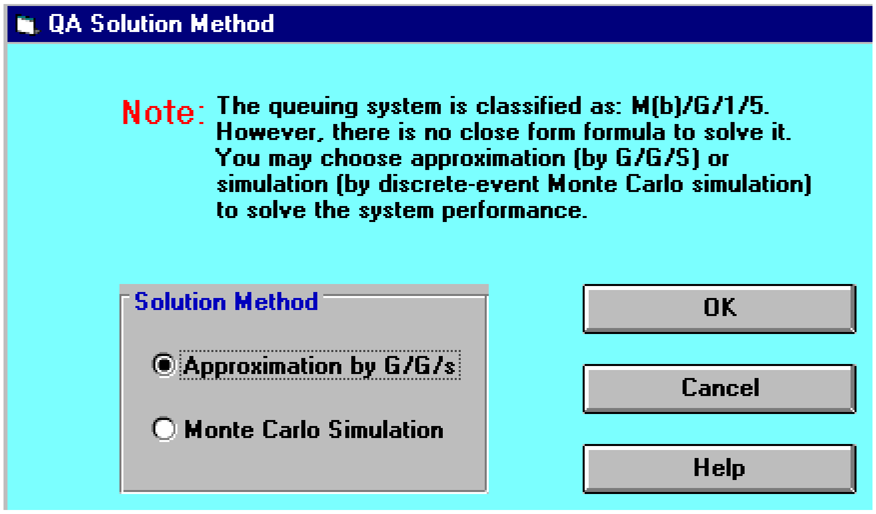
rate being *lamda* then the arrival rate changes from *lamda* to $(r/(n+1))^s$*lamda* when there are n customers in the system and n>=r.

- batch (bulk) size distribution - customers can arrive together (in batches, also known as in bulk) and this indicates the distribution of size of such batches.

As an indication of the analysis that can be done an example problem is shown below:

Data Description	ENTRY
Number of servers	1
Service time distribution (in hour)	Normal
Mean (u)	.7
Standard deviation (s>0)	.2
(Not used)	
Service pressure coefficient	1.5
Interarrival time distribution (in hour)	Exponential
Location parameter (a)	
Scale parameter (b>0) (b=mean if a=0)	.5
(Not used)	
Arrival discourage coefficient	1.7
Batch (bulk) size distribution	Normal
Mean (u)	3
Standard deviation (s>0)	0.5
(Not used)	
Queue capacity (maximum waiting space)	4
Customer population	
Busy server cost per hour	10
Idle server cost per hour	100
Customer waiting cost per hour	500
Customer being served cost per hour	5
Cost of customer being balked	600
Unit queue capacity cost	

Solving the problem we get:

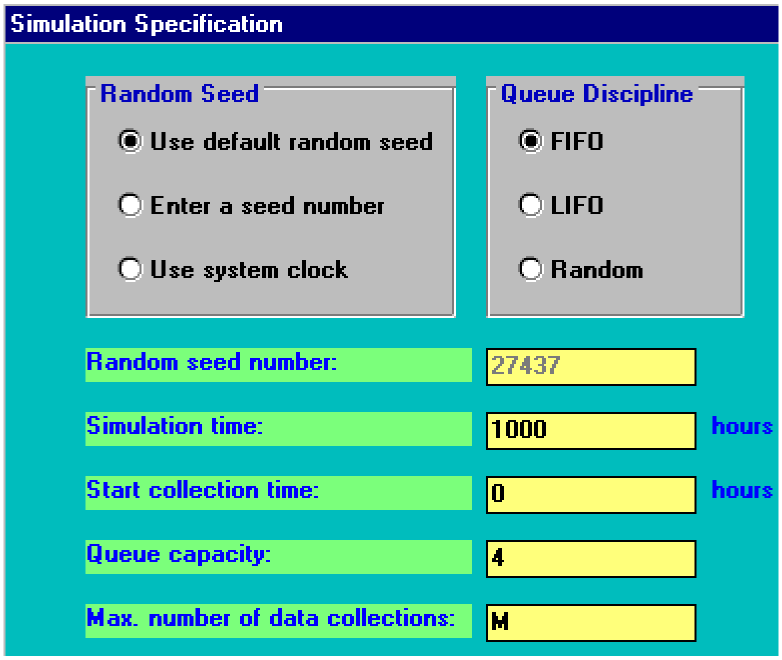


This screen indicates that no formulae exist to evaluate the situation we have set up. We can try to evaluate this situation using an approximation formula, or by Monte Carlo Simulation. If we choose to adopt the approximation approach we get:

11-14-2000	Performance Measure	Result
1	System: M(b)/G/1/5	From Approximation
2	Customer arrival rate (lambda) per hour =	2.0000
3	Service rate per server (mu) per hour =	1.4286
4	Overall system effective arrival rate per hour =	2.0000
5	Overall system effective service rate per hour =	2.0000
6	Overall system utilization =	420.0000 %
7	Average number of customers in the system (L) =	1.2187
8	Average number of customers in the queue (Lq) =	-2.9813
9	Average number of customers in the queue for a busy system (Lb) =	-0.7098
10	Average time customer spends in the system (W) =	0.6094 hours
11	Average time customer spends in the queue (Wq) =	-1.4906 hours
12	Average time customer spends in the queue for a busy system (Wb) =	-0.3549 hours
13	The probability that all servers are idle (Po) =	-320.0000 %
14	The probability an arriving customer waits (Pw or Pb) =	420.0000 %
15	Average number of customers being balked per hour =	0
16	Total cost of busy server per hour =	$42.0000
17	Total cost of idle server per hour =	$-320.0000
18	Total cost of customer waiting per hour =	$-1490.6250
19	Total cost of customer being served per hour =	$21.0000
20	Total cost of customer being balked per hour =	$0
21	Total queue space cost per hour =	$0
22	Total system cost per hour =	$-1747.6250

The difficulty is that these approximation results are plainly nonsense (i.e. not a good approximation). For example the average number of customers in the queue is -2.9813, the probability that all servers are idle is -320%, etc. Whilst for this particular case it is obvious that approximation (or perhaps the package) is not working, for other problems it may not be readily apparent that approximation does not work.

If we adopt the Monte Carlo Simulation approach then we have the screen below.



What will happen here is that the computer will construct a model of the system we have specified and internally generate customer arrivals, service times, etc and collect statistics on how the system performs. As specified above it will do this for 1000 time units (hours in this case). The phrase "Monte Carlo" derives from the well-known gambling city on the Mediterranean in Monaco. Just as in roulette we get random numbers produced by a roulette wheel when it is spun, so in Monte Carlo simulation we make use of random numbers generated by a computer.

The results are shown below:

11-14-2000	Performance Measure	Result
1	System: M(b)/G/1/5	From Simulation
2	Customer arrival rate (lambda) per hour =	2.0000
3	Service rate per server (mu) per hour =	1.4286
4	Overall system effective arrival rate per hour =	1.4129
5	Overall system effective service rate per hour =	1.4079
6	Overall system utilization =	99.7974 %
7	Average number of customers in the system (L) =	4.2882
8	Average number of customers in the queue (Lq) =	3.2902
9	Average number of customers in the queue for a busy system (Lb) =	3.2969
10	Average time customer spends in the system (W) =	3.0418 hours
11	Average time customer spends in the queue (Wq) =	2.3330 hours
12	Average time customer spends in the queue for a busy system (Wb) =	2.3377 hours
13	The probability that all servers are idle (Po) =	0.2026 %
14	The probability an arriving customer waits (Pw or Pb) =	99.7974 %
15	Average number of customers being balked per hour =	4.6478
16	Total cost of busy server per hour =	$9.9797
17	Total cost of idle server per hour =	$0.2025
18	Total cost of customer waiting per hour =	$1648.1810
19	Total cost of customer being served per hour =	$5.0071
20	Total cost of customer being balked per hour =	$2788.6530
21	Total queue space cost per hour =	$0
22	Total system cost per hour =	$4452.0230
23	Simulation time in hour =	1000.0000
24	Starting data collection time in hour =	0
25	Number of observations collected =	1408
26	Maximum number of customers in the queue =	4
27	Total simulation CPU time in second =	2.4170

These results seem much more reasonable than the results obtained the approximation.

However one factor to take into consideration is the simulation time we specified - here 1000 hours. In order to collect more accurate information on the behaviour of the system we might wish to simulate for longer. The results for simulating both 10 and 100 times as long are shown below.

11-14-2000	Performance Measure	Result
1	System: M(b)/G/1/5	From Simulation
2	Customer arrival rate (lambda) per hour =	2.0000
3	Service rate per server (mu) per hour =	1.4286
4	Overall system effective arrival rate per hour =	1.4272
5	Overall system effective service rate per hour =	1.4268
6	Overall system utilization =	99.8225 %
7	Average number of customers in the system (L) =	4.2827
8	Average number of customers in the queue (Lq) =	3.2844
9	Average number of customers in the queue for a busy system (Lb) =	3.2903
10	Average time customer spends in the system (W) =	3.0010 hours
11	Average time customer spends in the queue (Wq) =	2.3013 hours
12	Average time customer spends in the queue for a busy system (Wb) =	2.3054 hours
13	The probability that all servers are idle (Po) =	0.1775 %
14	The probability an arriving customer waits (Pw or Pb) =	99.8225 %
15	Average number of customers being balked per hour =	4.6027
16	Total cost of busy server per hour =	$9.9823
17	Total cost of idle server per hour =	$0.1766
18	Total cost of customer waiting per hour =	$1642.2850
19	Total cost of customer being served per hour =	$4.9925
20	Total cost of customer being balked per hour =	$2761.6240
21	Total queue space cost per hour =	$0
22	Total system cost per hour =	$4419.0610
23	Simulation time in hour =	10000.0000
24	Starting data collection time in hour =	0
25	Number of observations collected =	14269
26	Maximum number of customers in the queue =	4
27	Total simulation CPU time in second =	23.6630

11-14-2000	Performance Measure	Result
1	System: M(b)/G/1/5	From Simulation
2	Customer arrival rate (lambda) per hour =	2.0000
3	Service rate per server (mu) per hour =	1.4286
4	Overall system effective arrival rate per hour =	1.4266
5	Overall system effective service rate per hour =	1.4265
6	Overall system utilization =	99.7353 %
7	Average number of customers in the system (L) =	4.2723
8	Average number of customers in the queue (Lq) =	3.2751
9	Average number of customers in the queue for a busy system (Lb) =	3.2838
10	Average time customer spends in the system (W) =	2.9950 hours
11	Average time customer spends in the queue (Wq) =	2.2958 hours
12	Average time customer spends in the queue for a busy system (Wb) =	2.3019 hours
13	The probability that all servers are idle (Po) =	0.2647 %
14	The probability an arriving customer waits (Pw or Pb) =	99.7353 %
15	Average number of customers being balked per hour =	4.5840
16	Total cost of busy server per hour =	$9.9718
17	Total cost of idle server per hour =	$0.2816
18	Total cost of customer waiting per hour =	$1637.5800
19	Total cost of customer being served per hour =	$4.9870
20	Total cost of customer being balked per hour =	$2750.4130
21	Total queue space cost per hour =	$0
22	Total system cost per hour =	$4403.2330
23	Simulation time in hour =	100000.0000
24	Starting data collection time in hour =	0
25	Number of observations collected =	142654
26	Maximum number of customers in the queue =	4
27	Total simulation CPU time in second =	245.9250

Clearly the longer we simulate, the more confidence we may have in the statistics/probabilities obtained.

As before we can investigate how the system might behave with more servers. Simulating for 1000 hours (to reduce the overall elapsed time required) and looking at just the total system cost per hour (item 22 in the above outputs) we have the following:

```
Number of servers   Total system cost
       1                  4452
       2                  3314
       3                  2221
       4                  1614
       5                  1257
       6                   992
       7                   832
       8                   754
       9                   718
      10                   772
      11                   833
      12                   902
```

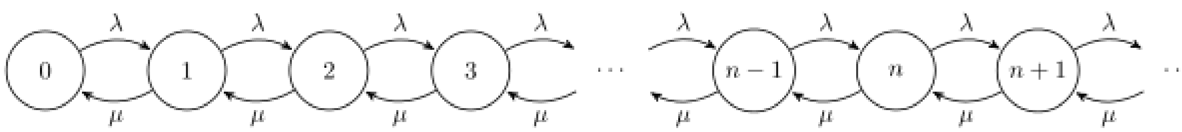
Hence here the number of servers associated with the minimum total system cost is 9

An M/M/1 queue is a stochastic process whose state space is the set {0,1,2,3,...} where the value corresponds to the number of customers in the system, including any currently in service.

- Arrivals occur at rate λ according to a Poisson process and move the process from state *i* to *i* + 1.
- Service times have an exponential distribution with rate parameter μ in the M/M/1 queue, where 1/μ is the mean service time.
- A single server serves customers one at a time from the front of the queue, according to a first-come, first-served discipline. When the service is complete the customer leaves the queue and the number of customers in the system reduces by one.
- The buffer is of infinite size, so there is no limit on the number of customers it can contain.

The model can be described as a continuous time Markov chain with transition rate matrix

on the state space {0,1,2,3,...}. This is the same continuous time Markov chain as in a birth–death process. The state space diagram for this chain is as below.



Transient solution

We can write a probability mass function dependent on *t* to describe the probability that the M/M/1 queue is in a particular state at a given time. We assume that the queue is initially in state *i* and write $p_k(t)$ for the probability of being in state *k* at time *t*. Then[2]

where , and is the modified Bessel function of the first kind. Moments for the transient solution can be expressed as the sum of two monotone functions.[3]

Stationary analysis

The model is considered stable only if λ < μ. If, on average, arrivals happen faster than service completions the queue will grow indefinitely long and the system will not have a stationary distribution. The stationary distribution is the limiting distribution for large values of *t*.

Various performance measures can be computed explicitly for the M/M/1 queue. We write ρ = λ/μ for the utilization of the buffer and require ρ < 1 for the queue to be stable. ρ represents the average proportion of time which the server is occupied.

Number of customers in the system

The probability that the stationary process is in state i (contains i customers, including those in service) is[4]:172–173

We see that the number of customers in the system is geometrically distributed with parameter $1 - \rho$. Thus the average number of customers in the system is $\rho/(1 - \rho)$ and the variance of number of customers in the system is $\rho/(1 - \rho)^2$. This result holds for any work conserving service regime, such as processor sharing.[5]

Busy period of server

The busy period is the time period measured between the instant a customer arrives to an empty system until the instant a customer departs leaving behind an empty system. The busy period has probability density function[6][7][8][9]

where I_1 is a modified Bessel function of the first kind,[10] obtained by using Laplace transforms and inverting the solution.[11]

The Laplace transform of the M/M/1 busy period is given by[12][13][14]:215

which gives the moments of the busy period, in particular the mean is $1/(\mu - \lambda)$ and variance is given by

Response time

The average response time or sojourn time (total time a customer spends in the system) does not depend on scheduling discipline and can be computed using Little's law as $1/(\mu - \lambda)$. The average time spent waiting is $1/(\mu - \lambda) - 1/\mu = \rho/(\mu - \lambda)$. The distribution of response times experienced does depend on scheduling discipline.

First-come, first-served discipline

For customers who arrive and find the queue as a stationary process, the response time they experience (the sum of both waiting time and service time) has transform $(\mu - \lambda)/(s + \mu - \lambda)$[15] and therefore probability density function[16]

Processor sharing discipline

In an M/M/1-PS queue there is no waiting line and all jobs receive an equal proportion of the service capacity.[17] Suppose the single server serves at rate 16 and there are 4 jobs in the system, each job will experience service at rate 4. The rate at which jobs receive service changes each time a job arrives at or departs from the system.[17]

For customers who arrive to find the queue as a stationary process, the Laplace transform of the distribution of response times experienced by customers was published in 1970,[17]for which an integral representation is known.[18] The waiting time distribution (response time less service time) for a customer requiring *x* amount of service has transform[4]:356

where *r* is the smaller root of the equation

The mean response time for a job arriving and requiring amount *x* of service can therefore be computed as *x* $\mu/(\mu - \lambda)$. An alternative approach computes the same results using a spectral expansion method.[5]

Markov Chain

In probability theory and related fields, a **Markov process**, named after the Russian mathematician Andrey Markov, is a stochastic process that satisfies the Markov property[1][2][3][4](sometimes characterized as "memorylessness"). Roughly speaking, a process satisfies the Markov property if one can make predictions for the future of the process based solely on its present state just as well as one could knowing the process's full history, hence independently from such history; i.e., conditional on the present state of the system, its future and past states are independent.

A **Markov chain** is a type of Markov process that has either discrete state space or discrete index set (often representing time), but the precise definition of a Markov chain varies.[5]For example, it is common to define a Markov chain as a Markov process in either discrete or continuous time with a countable state space (thus regardless of the nature of time),[6][7][8][9][10] but it is also common to define a Markov chain as having discrete time in either countable or continuous state space (thus regardless of the state space).

Markov studied Markov processes in the early 20th century, publishing his first paper on the topic in 1906.[6][11][12][13] Random walks on integers and the gambler's ruin problem are examples of Markov processes. [14][15][16] Some variations of these processes were studied hundreds of years earlier in the context of independent variables.[17][18] Two important examples of Markov processes are the Wiener process, also known as the Brownian motion process, and the Poisson process,[19] which are considered the most important and central stochastic processes in the theory of stochastic processes,[20][21][22] and were discovered repeatedly and independently, both before and after 1906, in various settings.[23][24]These two processes are Markov processes in continuous time, while random walks on the integers and the gambler's ruin problem are examples of Markov processes in discrete time.

Markov chains have many applications as statistical models of real-world processes,[25][26][27] such as studying cruise control systems in motor vehicles, queues or lines of customers arriving at an airport, exchange rates of currencies, storage systems such as dams, and population growths of certain animal species.[28] The algorithm known as PageRank, which was originally proposed for the internet search engine Google, is based on a Markov process.[29][30] Furthermore, Markov processes are the basis for general stochastic simulation methods known as Gibbs sampling and Markov Chain Monte Carlo, are used for simulating random objects with specific probability distributions, and have found extensive application in Bayesian statistics.

A **Markov chain** is a mathematical system that experiences transitions from one state to another according to certain probabilistic rules. The defining characteristic of a Markov chain is that no matter *how* the process arrived at its present state, the possible future states are fixed. In other words, the probability of transitioning to any particular state is dependent solely on the current state and time elapsed. The **state space**, or set of all possible states, can be anything: letters, numbers, weather conditions, baseball scores, or stock performances.

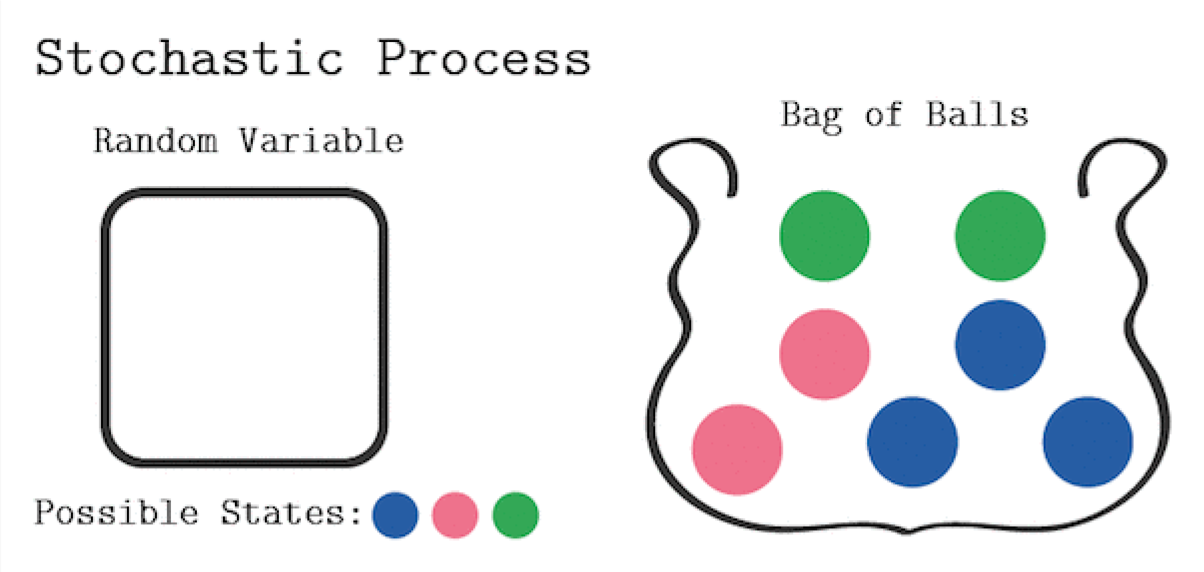
Markov chains may be modeled by finite state machines, and random walks provide a prolific example of their usefulness in mathematics. They arise broadly in statistical and information-theoretical contexts and are widely employed in economics, game theory, queueing (communication) theory, genetics, and finance. While it is possible to discuss Markov chains with any size of state space, the initial theory and most applications are focused on cases with a finite (or countably infinite) number of states.

Many uses of Markov chains require proficiency with common matrix methods.

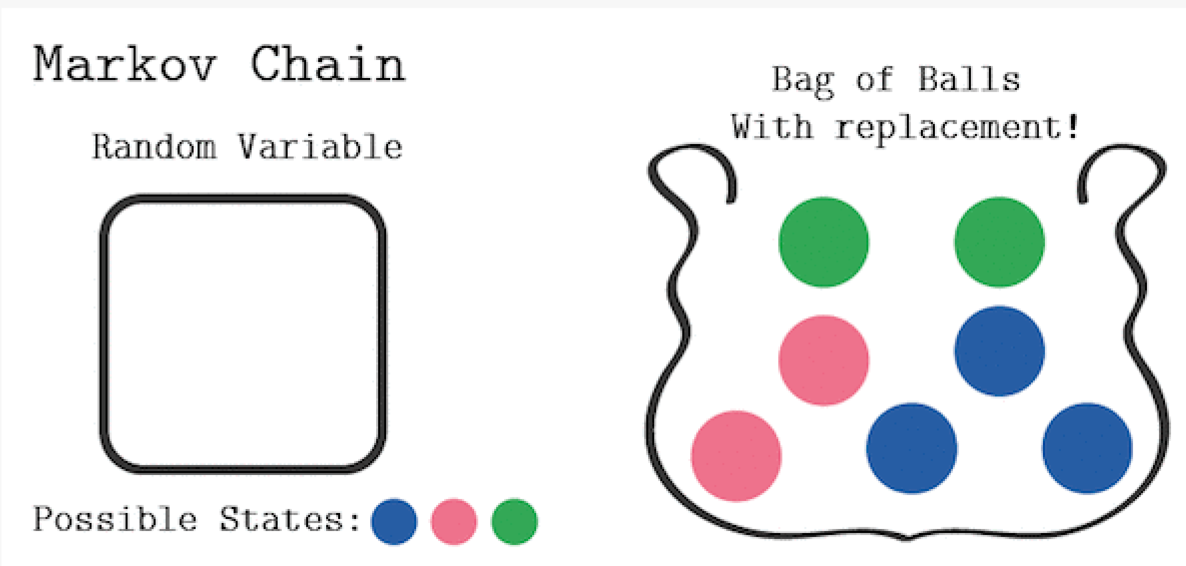
Basic Concept

A Markov chain is a stochastic process, but it differs from a general stochastic process in that a Markov chain must be "memory-less". That is, (the probability of) future actions are not dependent upon the steps that led up to the present state. This is called the **Markov property**. While the theory of Markov chains is important precisely because so many "everyday" processes satisfy the Markov property, there are many common examples of stochastic properties that do not satisfy the Markov property.

A common probability question asks what is the probability of getting a certain color ball, when selecting uniformly and at random from a bag of multicolored balls. It could also ask what the probability of the next ball is, and so on. In such a way, a stochastic process begins to exist with color for the random variable, and it does not satisfy the Markov property. Depending upon which balls are removed, the probability of getting a certain color ball later may be drastically different.

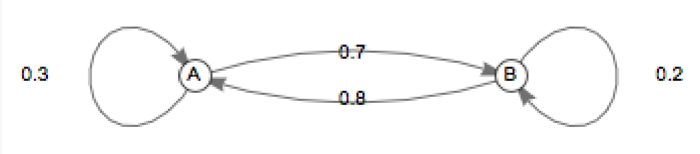


A variant of the same question asks once again for ball color, but it allows replacement each time a ball is drawn. Once again, this creates a stochastic process with color for the random variable. This process, however, **does** satisfy the Markov property. Can you figure out why?



In probability theory, the most immediate example is that of a *time-homogeneous Markov chain*, in which the probability of any state transition is independent of time. Such a process may be visualized with a labeled directed graph, for which the sum of the labels of any vertex's outgoing edges is 1.

A (time-homogeneous) Markov chain built on states A and B is depicted in the diagram below. What is the probability that a process beginning on A will be on B after 2 moves?



In order to move from A to B, the process must either stay on A the first move, then move to B the second move; or move to B the first move, then stay on B the second move. According to the diagram, the probability of that is

Alternatively, the probability that the process will be on A after 2 moves is . Since there are only two states in the chain, the process must be on B if it is not on A, and therefore, the probability that the process will be on B after 2 moves is

In the language of conditional probability and random variables, a Markov chain is a sequence of random variables satisfying the rule of conditional independence:

The Markov Property.

For any positive integer and possible states of the random variables,

In other words, knowledge of the previous state is all that is necessary to determine the probability distribution of the current state. This definition is broader than the one explored above, as it allows for *non-stationary transition probabilities* and therefore *time-inhomogeneous Markov chains*; that is, as time goes on (steps increase), the probability of moving from one state to another may change.

Transition Matrices

A **transition matrix** for Markov chain at time is a matrix containing information on the probability of transitioning between states. In particular, given an ordering of a matrix's rows and columns by the state space , the element of the matrix is given by

This means each row of the matrix is a *probability vector*, and the sum of its entries is .

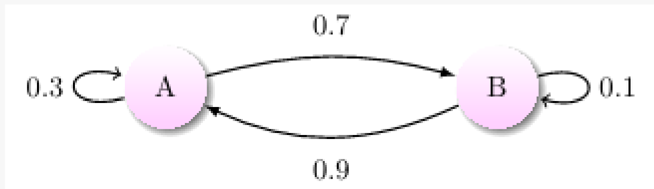
Transition matrices have the property that the product of subsequent ones describes a transition along the time interval spanned by the transition matrices. That is to say, has in its position the probability that given that . And, in general, the position of is the probability .

Prove that, for any natural number and states , the matrix entry .

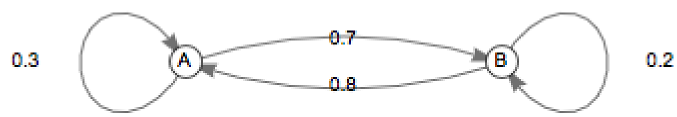
Denote . By matrix multiplication,The final equality follows from conditional probability.

The **-step** transition matrix is and, by the above, satisfies

For the time-independent Markov chain described by the picture below, what is its -step transition matrix?

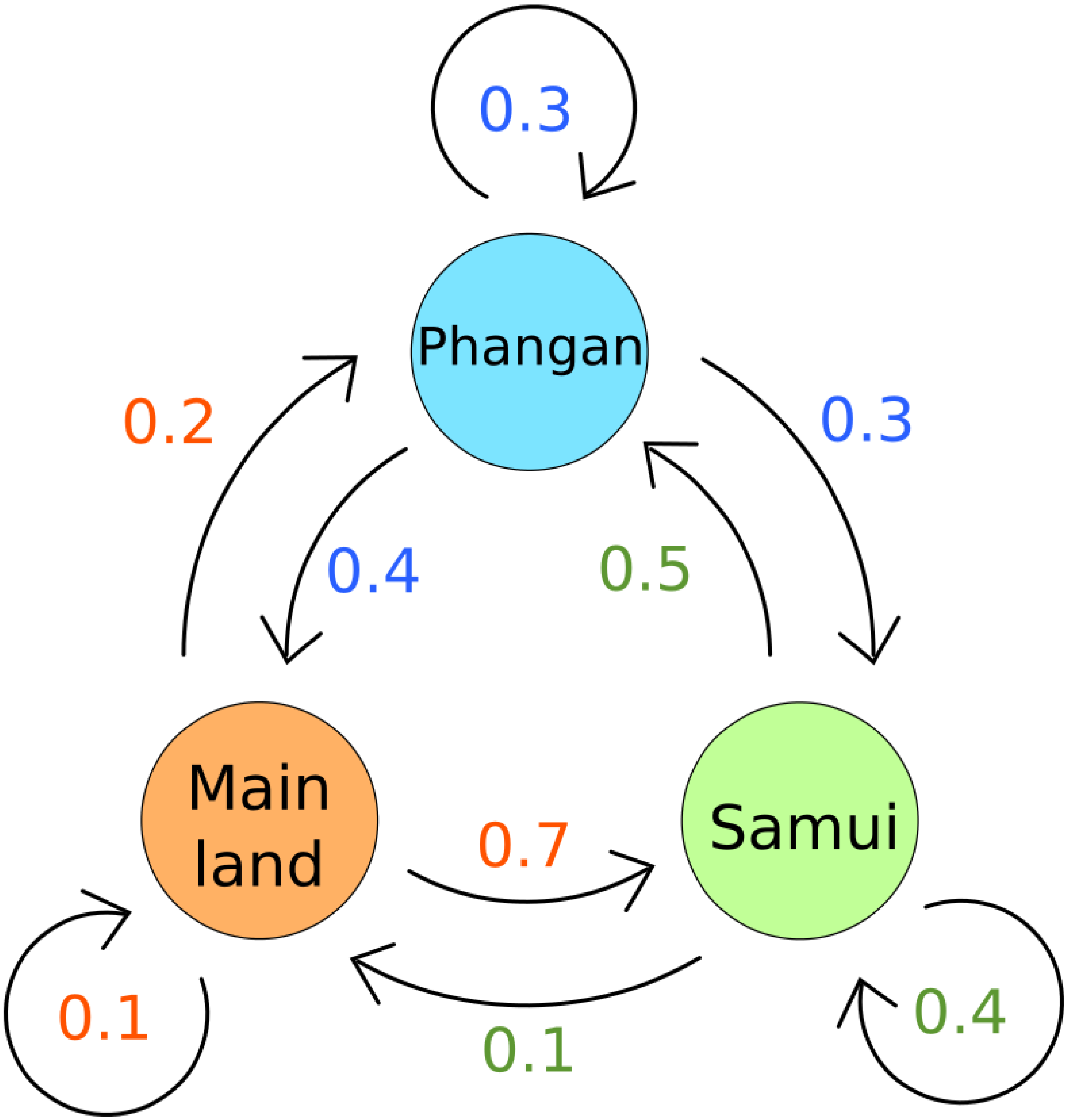


Note the transition matrix is, It follows that the -step transition matrix is
A Markov chain has first state A and second state B, and its transition probabilities for all time are given by the following graph.



What is its transition matrix?

Note: Note: The transition matrix is oriented such that the th **row** represents the set of probabilities of transitioning from state to another state.
Submit your answer



Once people arrive in Thailand, they want to enjoy the sun and beaches on 2 popular islands in the south: Samui Island &Phangan Island.

From survey data, when on the mainland, 70% of tourists plan to go to Samui Island, 20% to Phangan Island, and only 10% remain on shore the next day.

When on Samui Island, 40% continue to stay on Samui, 50% plan to go to Phangan Island, and only 10% return to mainland the next day.

Finally, when on Phangan Island, 30% prolong their stay here, 30% divert to Samui Island, and 40% go back to mainland the next day.

Starting from the mainland, what is the probability (in percentage) that the travelers will be on the mainland at the end of a 3-day trip?

Properties

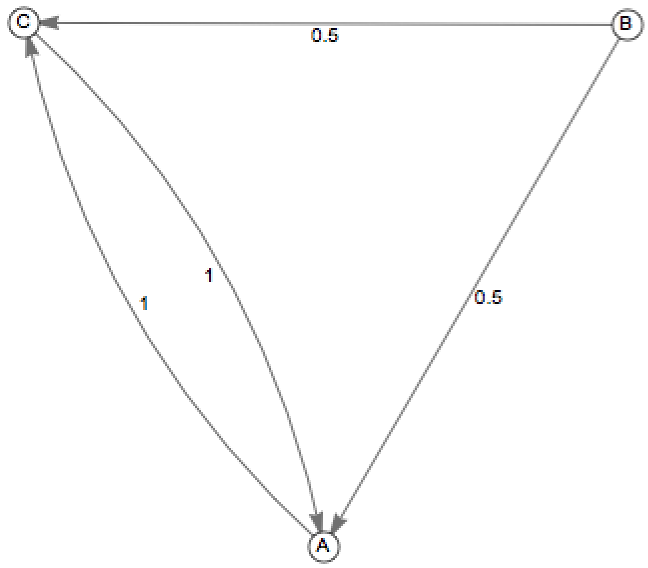
A variety of descriptions of either a specific state in a Markov chain or the entire Markov chain allow for better understanding of the Markov's chain behavior. Let be the transition matrix of Markov chain .

- A state has **period** if any chain starting at and returning to state with positive probability must take a number of steps divisible by . If , then the state is known as **aperiodic**, and if , the state is known as **periodic**. If all states are aperiodic, then the Markov chain is known as aperiodic.
- A Markov chain is known as **irreducible** if there exists a chain of steps between any two states that has positive probability.
- An **absorbing** state is a state for which . Absorbing states are crucial for the discussion of absorbing Markov chains.
- A state is known as recurrent or transient depending upon whether or not the Markov chain will eventually return to it. A recurrent state is known as positive recurrent if it is expected to return within a finite number of steps and null recurrent otherwise.
- A state is known as ergodic if it is positive recurrent and aperiodic. A Markov chain is ergodic if all its states are.

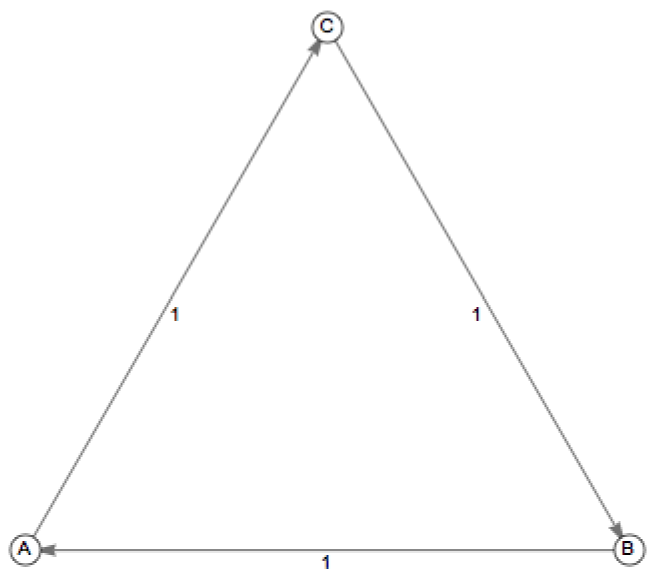
Irreducibility and periodicity both concern the locations a Markov chain could be at some later point in time, given where it started. Stationary distributions deal with the likelihood of a process being in a certain state at an unknown point of time. For Markov chains with a finite number of states, each of which is positive recurrent, an aperiodic Markov chain is the same as an irreducible Markov chain.

Both Markov chains are periodic and irreducibleNeither Markov chain is irreducibleChain 1 is aperiodic and irreducible, and chain 2 is aperiodicChain 1 is aperiodic, and chain 2 is irreducibleChain 1 is not irreducible, and chain 2 is not aperiodic

The graphs of two time-homogeneous Markov chains are shown below.



Chain 1



Chain 2

Determine facts about their periodicity and reducibility.

Before the Monte Carlo method was developed, simulations tested a previously understood deterministic problem and statistical sampling was used to estimate uncertainties in the simulations. Monte Carlo simulations invert this approach, solving deterministic problems using a probabilistic analog (see Simulated annealing).

An early variant of the Monte Carlo method can be seen in the Buffon's needle experiment, in which π can be estimated by dropping needles on a floor made of parallel and equidistant strips. In the 1930s, Enrico Fermi first experimented with the Monte Carlo method while studying neutron diffusion, but did not publish anything on it.[12]

The modern version of the Markov Chain Monte Carlo method was invented in the late 1940s by Stanislaw Ulam, while he was working on nuclear weapons projects at the Los Alamos National Laboratory. Immediately after Ulam's breakthrough, John von Neumann understood its importance and programmed the ENIAC computer to carry out Monte Carlo calculations. In 1946, physicists at Los Alamos Scientific Laboratory were investigating radiation shielding and the distance that neutrons would likely travel through various materials. Despite having most of the necessary data, such as the average distance a neutron would travel in a substance before it collided with an atomic nucleus, and how much energy the neutron was likely to give off following a collision, the Los Alamos physicists were unable to solve the problem using conventional, deterministic mathematical methods. Ulam had the idea of using random experiments. He recounts his inspiration as follows:

> The first thoughts and attempts I made to practice [the Monte Carlo Method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later [in 1946], I described the idea to John von Neumann, and we began to plan actual calculations.[13]

Being secret, the work of von Neumann and Ulam required a code name.[citation needed] A colleague of von Neumann and Ulam, Nicholas Metropolis, suggested using the name *Monte Carlo*, which refers to the Monte Carlo Casino in Monaco where Ulam's uncle would borrow money from relatives to gamble.[12] Using lists of "truly random" random numbers was extremely slow, but von Neumann developed a way to calculate pseudorandom numbers, using the middle-square method. Though this method has been criticized as crude, von Neumann was aware of this: he justified it as being faster than any other method at his disposal, and also noted that when it went awry it did so obviously, unlike methods that could be subtly incorrect.

Monte Carlo methods were central to the simulations required for the Manhattan Project, though severely limited by the computational tools at the time. In the 1950s they were used at Los Alamos for early work relating to the development of the hydrogen bomb, and became popularized in the fields of physics, physical chemistry, and operations research. The Rand Corporation and the U.S. Air Force were two of the major organizations responsible for funding and disseminating information on Monte Carlo methods during this time, and they began to find a wide application in many different fields.

The theory of more sophisticated mean field type particle Monte Carlo methods had certainly started by the mid-1960s, with the work of Henry P. McKean Jr. on Markov interpretations of a class of nonlinear parabolic partial differential equations arising in fluid mechanics.[14][15] We also quote an earlier pioneering article by Theodore E. Harris and Herman Kahn, published in 1951, using mean field genetic-type Monte Carlo methods for estimating particle transmission energies.[16]Mean field genetic type Monte Carlo methodologies are also used as heuristic natural search algorithms (a.k.a. Metaheuristic) in evolutionary computing. The origins of these mean field computational techniques can be traced to 1950 and 1954 with the work of Alan Turing on genetic type mutation-selection learning machines[17] and the articles by Nils AallBarricelli at the Institute for Advanced Study in Princeton, New Jersey.[18][19]

Quantum Monte Carlo, and more specifically Diffusion Monte Carlo methods can also be interpreted as a mean field particle Monte Carlo approximation of Feynman-Kac path integrals.[20][21][22][23][24][25][26] The origins of Quantum Monte Carlo methods are often attributed to Enrico Fermi and Robert Richtmyer who developed in 1948 a mean field particle interpretation of neutron-chain reactions,[27] but the first heuristic-like and genetic type particle algorithm (a.k.a. Resampled or Reconfiguration Monte Carlo methods) for estimating ground state energies of quantum systems (in reduced matrix models) is due to Jack H. Hetherington in 1984[26] In molecular chemistry, the use of genetic heuristic-like particle methodologies (a.k.a. pruning and enrichment strategies) can be traced back to 1955 with the seminal work of Marshall. N. Rosenbluth and Arianna. W. Rosenbluth.[28]

The use of Sequential Monte Carlo in advanced signal processing and Bayesian inference is more recent. It was in 1993, that Gordon et al., published in their seminal work[29] the first application of a Monte Carlo resampling algorithm in Bayesian statistical inference. The authors named their algorithm 'the bootstrap filter', and demonstrated that compared to other filtering methods, their bootstrap algorithm does not require any assumption about that state-space or the noise of the system. We also quote another pioneering article in this field of Genshiro Kitagawa on a related "Monte Carlo filter",[30] and the ones by Pierre Del Moral[31]and HimilconCarvalho, Pierre Del Moral, André Monin and Gérard Salut[32] on particle filters published in the mid-1990s. Particle filters were also developed in signal processing in the early 1989-1992 by P. Del Moral, J.C. Noyer, G. Rigal, and G. Salut in the LAAS-CNRS in a series of restricted and classified research reports with STCAN (Service Technique des Constructions et ArmesNavales), the IT company DIGILOG, and the LAAS-CNRS (the Laboratory for Analysis and Architecture of Systems) on RADAR/SONAR and GPS signal processing problems.[33][34][35][36][37][38] These Sequential Monte Carlo methodologies can be interpreted as an acceptance-rejection sampler equipped with an interacting recycling mechanism.

From 1950 to 1996, all the publications on Sequential Monte Carlo methodologies including the pruning and resample Monte Carlo methods introduced in computational physics and molecular chemistry, present natural and heuristic-like algorithms applied to different situations without a single proof of their consistency, nor a discussion on the bias of the estimates and on genealogical and ancestral tree based algorithms. The mathematical foundations and the first rigorous analysis of these particle algorithms are due to Pierre Del Moral[31][39] in 1996. Branching type particle methodologies with varying population sizes were also developed in the end of the 1990s by Dan Crisan, Jessica Gaines and Terry Lyons,[40][41][42] and by Dan Crisan, Pierre Del Moral and Terry Lyons.[43Further developments in this field were developed in 2000 by P. Del Moral, A. Guionnet and L. Miclo.

Definitions

There is no consensus on how *Monte Carlo* should be defined. For example, Ripley[46] defines most probabilistic modeling as *stochastic simulation*, with *Monte Carlo* being reserved for Monte Carlo integration and Monte Carlo statistical tests. Sawilowsky[47] distinguishes between a simulation, a Monte Carlo method, and a Monte Carlo simulation: a simulation is a fictitious representation of reality, a Monte Carlo method is a technique that can be used to solve a mathematical or statistical problem, and a Monte Carlo simulation uses repeated sampling to determine the properties of some phenomenon (or behavior). Examples:

- Simulation: Drawing **one** pseudo-random uniform variable from the interval [0,1] can be used to simulate the tossing of a coin: If the value is less than or equal to 0.50 designate the outcome as heads, but if the value is greater than 0.50 designate the outcome as tails. This is a simulation, but not a Monte Carlo simulation.
- Monte Carlo method: Pouring out a box of coins on a table, and then computing the ratio of coins that land heads versus tails is a Monte Carlo method of determining the behavior of repeated coin tosses, but it is not a simulation.
- Monte Carlo simulation: Drawing **a large number** of pseudo-random uniform variables from the interval [0,1], and assigning values less than or equal to 0.50 as heads and greater than 0.50 as tails, is a *Monte Carlo simulation* of the behavior of repeatedly tossing a coin.

Kalos and Whitlock[11] point out that such distinctions are not always easy to maintain. For example, the emission of radiation from atoms is a natural stochastic process. It can be simulated directly, or its average behavior can be described by stochastic equations that can themselves be solved using Monte Carlo methods. "Indeed, the same computer code can be viewed simultaneously as a 'natural simulation' or as a solution of the equations by natural sampling."

Monte Carlo and random numbers

The main idea behind this method is that the results are computed based on repeated random sampling and statistical analysis. The Monte Carlo simulation is in fact random experimentations, in the case that, the results of these experiments are not well known. Monte Carlo simulations are typically characterized by a large number of unknown parameters, many of which are difficult to obtain experimentally.[48] Monte Carlo simulation methods do not always require truly random numbers to be useful (although, for some applications such as primality testing, unpredictability is vital).[49] Many of the most useful techniques use deterministic, pseudorandom sequences, making it easy to test and re-run simulations. The only quality usually necessary to make good simulations is for the pseudo-random sequence to appear "random enough" in a certain sense.

What this means depends on the application, but typically they should pass a series of statistical tests. Testing that the numbers are uniformly distributed or follow another desired distribution when a large enough number of elements of the sequence are considered is one of the simplest, and most common ones. Weak correlations between successive samples is also often desirable/necessary.

Sawilowsky lists the characteristics of a high quality Monte Carlo simulation:[47]

- the (pseudo-random) number generator has certain characteristics (*e.g.*, a long "period" before the sequence repeats)
- the (pseudo-random) number generator produces values that pass tests for randomness
- there are enough samples to ensure accurate results
- the proper sampling technique is used
- the algorithm used is valid for what is being modeled
- it simulates the phenomenon in question.

Pseudo-random number sampling algorithms are used to transform uniformly distributed pseudo-random numbers into numbers that are distributed according to a given probability distribution.

Low-discrepancy sequences are often used instead of random sampling from a space as they ensure even coverage and normally have a faster order of convergence than Monte Carlo simulations using random or pseudorandom sequences. Methods based on their use are called quasi-Monte Carlo methods.

In an effort to assess the impact of random number quality on Monte Carlo simulation outcomes, astrophysical researchers tested cryptographically-secure pseudorandom numbers generated via Intel's RdRand instruction set, as compared to those derived from algorithms, like the Mersenne Twister, in Monte Carlo simulations of radio flares from brown dwarfs. RdRand is the closest pseudorandom number generator to a true random number generator. No statistically-significant difference was found between models generated with typical pseudorandom number generators and RdRand for trials consisting of the generation of 10^7 random numbers.

Monte Carlo simulation versus "what if" scenarios

There are ways of using probabilities that are definitely not Monte Carlo simulations — for example, deterministic modeling using single-point estimates. Each uncertain variable within a model is assigned a “best guess” estimate. Scenarios (such as best, worst, or most likely case) for each input variable are chosen and the results recorded.[51]

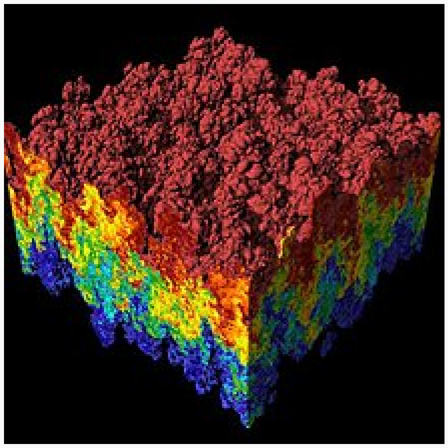
By contrast, Monte Carlo simulations sample from a probability distribution for each variable to produce hundreds or thousands of possible outcomes. The results are analyzed to get probabilities of different outcomes occurring.[52] For example, a comparison of a spreadsheet cost construction model run using traditional “what if” scenarios, and then running the comparison again with Monte Carlo simulation and triangular probability distributions shows that the Monte Carlo analysis has a narrower range than the “what if” analysis.[examples needed] This is because the “what if” analysis gives equal weight to all scenarios (see quantifying uncertainty in corporate finance), while the Monte Carlo method hardly samples in the very low probability regions. The samples in such regions are called "rare events".

Applications

Monte Carlo methods are especially useful for simulating phenomena with significant uncertainty in inputs and systems with a large number of coupled degrees of freedom. Areas of application include:

Physical sciences

Computational physics



Numerical analysis · Simulation

Data analysis · Visualization

Potentials[show]

Fluid dynamics[show]

Monte Carlo methods[show]

Particle[show]

Scientists[show]

- v
- t
- e

See also: Monte Carlo method in statistical physics

Monte Carlo methods are very important in computational physics, physical chemistry, and related applied fields, and have diverse applications from complicated quantum chromodynamics calculations to designing heat shieldsand aerodynamic forms as well as in modeling radiation transport for radiation dosimetry calculations.[53][54][55] In statistical physics Monte Carlo molecular modeling is an alternative to computational molecular dynamics, and Monte Carlo methods are used to compute statistical field theories of simple particle and polymer systems.[28][56]Quantum Monte Carlo methods solve the many-body problem for quantum systems.[8][9][20] In radiation materials science, the binary collision approximation for simulating ion implantation is usually based on a Monte Carlo approach to select the next colliding atom.[57] In experimental particle physics, Monte Carlo methods are used for designing detectors, understanding their behavior and comparing experimental data to theory. In astrophysics, they are used in such diverse manners as to model both galaxy evolution[58] and microwave radiation transmission through a rough planetary surface.[59] Monte Carlo methods are also used in the ensemble models that form the basis of modern weather forecasting.

Engineering

Monte Carlo methods are widely used in engineering for sensitivity analysis and quantitative probabilistic analysis in process design. The need arises from the interactive, co-linear and non-linear behavior of typical process simulations. For example,

- In microelectronics engineering, Monte Carlo methods are applied to analyze correlated and uncorrelated variations in analog and digital integrated circuits.
- In geostatistics and geometallurgy, Monte Carlo methods underpin the design of mineral processing flowsheetsand contribute to quantitative risk analysis.
- In wind energy yield analysis, the predicted energy output of a wind farm during its lifetime is calculated giving different levels of uncertainty (P90, P50, etc.)
- impacts of pollution are simulated[60] and diesel compared with petrol.[61]
- In fluid dynamics, in particular rarefied gas dynamics, where the Boltzmann equation is solved for finite Knudsen number fluid flows using the direct simulation Monte Carlo [62] method in combination with highly efficient computational algorithms.[63]
- In autonomous robotics, Monte Carlo localization can determine the position of a robot. It is often applied to stochastic filters such as the Kalman filter or particle filter that forms the heart of the SLAM (simultaneous localization and mapping) algorithm.
- In telecommunications, when planning a wireless network, design must be proved to work for a wide variety of scenarios that depend mainly on the number of users, their locations and the services they want to use. Monte Carlo methods are typically used to generate these users and their states. The network performance is then evaluated and, if results are not satisfactory, the network design goes through an optimization process.
- In reliability engineering, one can use Monte Carlo simulation to generate mean time between failures and mean time to repair for components.
- In signal processing and Bayesian inference, particle filters and sequential Monte Carlo techniques are a class of mean field particle methods for sampling and computing the posterior distribution of a signal process given some noisy and partial observations using interacting empirical measures.

Climate change and radiative forcing

The Intergovernmental Panel on Climate Change relies on Monte Carlo methods in probability density function analysis of radiative forcing.

Probability density function (PDF) of ERF due to total GHG, aerosol forcing and total anthropogenic forcing. The GHG consists of WMGHG, ozone and stratospheric water vapour. The PDFs are generated based on uncertainties provided in Table 8.6. The combination of the individual RF agents to derive total forcing over the Industrial Era are done by Monte Carlo simulations and based on the method in Boucher and Haywood (2001). PDF of the ERF from surface albedo changes and combined contrails and contrail-induced cirrus are included in the total anthropogenic forcing, but not shown as a separate PDF. We currently do not have ERF estimates for some forcing mechanisms: ozone, land use, solar, etc.

Computational biology

Monte Carlo methods are used in various fields of computational biology, for example for Bayesian inference in phylogeny, or for studying biological systems such as genomes, proteins,[65] or membranes.[66] The systems can be studied in the coarse-grained or *ab initio* frameworks depending on the desired accuracy. Computer simulations allow us to monitor the local environment of a particular molecule to see if some chemical reaction is happening for instance. In cases where it is not feasible to conduct a physical experiment, thought experiments can be conducted (for instance: breaking bonds, introducing impurities at specific sites, changing the local/global structure, or introducing external fields).

Computer graphics

Path tracing, occasionally referred to as Monte Carlo ray tracing, renders a 3D scene by randomly tracing samples of possible light paths. Repeated sampling of any given pixel will eventually cause the average of the samples to converge on the correct solution of the rendering equation, making it one of the most physically accurate 3D graphics rendering methods in existence.

Applied statistics

The standards for Monte Carlo experiments in statistics were set by Sawilowsky.[67][68] In applied statistics, Monte Carlo methods are generally used for three purposes:

1. To compare competing statistics for small samples under realistic data conditions. Although type I error and power properties of statistics can be calculated for data drawn from classical theoretical distributions (*e.g.*, normal curve, Cauchy distribution) for asymptotic conditions (*i. e*, infinite sample size and infinitesimally small treatment effect), real data often do not have such distributions.[69]
2. To provide implementations of hypothesis tests that are more efficient than exact tests such as permutation tests (which are often impossible to compute) while being more accurate than critical values for asymptotic distributions.
3. To provide a random sample from the posterior distribution in Bayesian inference. This sample then approximates and summarizes all the essential features of the posterior.

Monte Carlo methods are also a compromise between approximate randomization and permutation tests. An approximate randomization test is based on a specified subset of all permutations (which entails potentially enormous housekeeping of which permutations have been considered). The Monte Carlo approach is based on a specified number of randomly drawn permutations (exchanging a minor loss in precision if a permutation is drawn twice – or more frequently—for the efficiency of not having to track which permutations have already been selected).

Artificial intelligence for games

Main article: Monte Carlo tree search

Monte Carlo methods have been developed into a technique called Monte-Carlo tree search that is useful for searching for the best move in a game. Possible moves are organized in a search tree and a large number of random simulations are used to estimate the long-term potential of each move. A black box simulator represents the opponent's moves.[70]

The Monte Carlo tree search (MCTS) method has four steps:[71]

1. Starting at root node of the tree, select optimal child nodes until a leaf node is reached.
2. Expand the leaf node and choose one of its children.
3. Play a simulated game starting with that node.
4. Use the results of that simulated game to update the node and its ancestors.

The net effect, over the course of many simulated games, is that the value of a node representing a move will go up or down, hopefully corresponding to whether or not that node represents a good move.

Monte Carlo Tree Search has been used successfully to play games such as Go, Tantrix, Battleship, Havannah, and Arimaa.

See also: Computer Go

Design and visuals

Monte Carlo methods are also efficient in solving coupled integral differential equations of radiation fields and energy transport, and thus these methods have been used in global illumination computations that produce photo-realistic images of virtual 3D models, with applications in video games, architecture, design, computer generated films, and cinematic special effects.

Search and rescue

The US Coast Guard utilizes Monte Carlo methods within its computer modeling software SAROPS in order to calculate the probable locations of vessels during search and rescue operations. Each simulation can generate as many as ten thousand data points which are randomly distributed based upon provided variables.[78] Search patterns are then generated based upon extrapolations of these data in order to optimize the probability of containment (POC) and the probability of detection (POD), which together will equal an overall probability of success (POS). Ultimately this serves as a practical application of probability distribution in order to provide the swiftest and most expedient method of rescue, saving both lives and resources.

Finance and business

See also: Monte Carlo methods in finance, Quasi-Monte Carlo methods in finance, Monte Carlo methods for option pricing, Stochastic modelling (insurance), and Stochastic asset model

Monte Carlo simulation is commonly used to evaluate the risk and uncertainty that would affect the outcome of different decision options. Typically, this is achieved using spreadsheet risk analysis add-ins. Monte Carlo simulation allows the business risk analyst to incorporate the total effects of uncertainty in variables like sales volume, commodity and labour prices, interest and exchange rates, as well as the effect of distinct risk events like the cancellation of a contract or the change of a tax law.

Monte Carlo methods in finance are often used to evaluate investments in projects at a business unit or corporate level, or to evaluate financial derivatives. They can be used to model project schedules, where simulations aggregate estimates for worst-case, best-case, and most likely durations for each task to determine outcomes for the overall project. Monte Carlo methods are also used in option pricing, default risk analysis.

Law

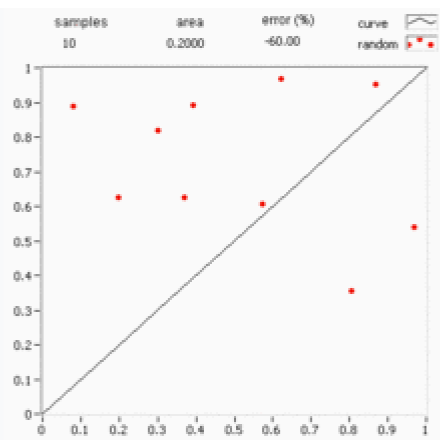
A Monte Carlo approach was used for evaluating the potential value of a proposed program to help female petitioners in Wisconsin be successful in their applications for harassment and domestic abuse restraining orders. It was proposed to help women succeed in their petitions by providing them with greater advocacy thereby potentially reducing the risk of rape and physical assault. However, there were many variables in play that could not be estimated perfectly, including the effectiveness of restraining orders, the success rate of petitioners both with and without advocacy, and many others. The study ran trials which varied these variables to come up with an overall estimate of the success level of the proposed program as a whole.

Use in mathematics

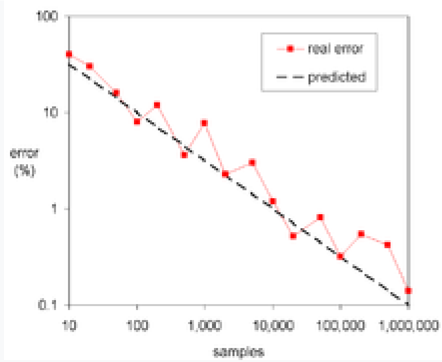
In general, Monte Carlo methods are used in mathematics to solve various problems by generating suitable random numbers (see also Random number generation) and observing that fraction of the numbers that obeys some property or properties. The method is useful for obtaining numerical solutions to problems too complicated to solve analytically. The most common application of the Monte Carlo method is Monte Carlo integration.

Integration

Main article: Monte Carlo integration



Monte-Carlo integration works by comparing random points with the value of the function



Errors reduce by a factor of

Deterministic numerical integration algorithms work well in a small number of dimensions, but encounter two problems when the functions have many variables. First, the number of function evaluations needed increases rapidly with the number of dimensions. For example, if 10 evaluations provide adequate accuracy in one dimension, then 10^{100} points are needed for 100 dimensions—far too many to be computed. This is called the curse of dimensionality. Second, the boundary of a multidimensional region may be very complicated, so it may not be feasible to reduce the problem to an iterated integral.[84] 100 dimensions is by no means unusual, since in many physical problems, a "dimension" is equivalent to a degree of freedom.

Monte Carlo methods provide a way out of this exponential increase in computation time. As long as the function in question is reasonably well-behaved, it can be estimated by randomly selecting points in 100-dimensional space, and taking some kind of average of the function values at these points. By the central limit theorem, this method displays convergence—i.e., quadrupling the number of sampled points halves the error, regardless of the number of dimensions.[84]

A refinement of this method, known as importance sampling in statistics, involves sampling the points randomly, but more frequently where the integrand is large. To do this precisely one would have to already know the integral, but one can approximate the integral by an integral of a similar function or use adaptive routines such as stratified sampling, recursive stratified sampling, adaptive umbrella sampling[85][86] or the VEGAS algorithm.

A similar approach, the quasi-Monte Carlo method, uses low-discrepancy sequences. These sequences "fill" the area better and sample the most important points more frequently, so quasi-Monte Carlo methods can often converge on the integral more quickly.

Another class of methods for sampling points in a volume is to simulate random walks over it (Markov chain Monte Carlo). Such methods include the Metropolis-Hastings algorithm, Gibbs sampling, Wang and Landau algorithm, and interacting type MCMC methodologies such as the sequential Monte Carlo samplers.

Simulation and optimization

Another powerful and very popular application for random numbers in numerical simulation is in numerical optimization. The problem is to minimize (or maximize) functions of some vector that often has a large number of dimensions. Many problems can be phrased in this way: for example, a computer chess program could be seen as trying to find the set of, say, 10 moves that produces the best evaluation function at the end. In the traveling salesman problem the goal is to minimize distance traveled. There are also applications to engineering design, such as multidisciplinary design optimization. It has been applied with quasi-one-dimensional models to solve particle dynamics problems by efficiently exploring large configuration space. Reference [88] is a comprehensive review of many issues related to simulation and optimization.

The traveling salesman problem is what is called a conventional optimization problem. That is, all the facts (distances between each destination point) needed to determine the optimal path to follow are known with certainty and the goal is to run through the possible travel choices to come up with the one with the lowest total distance. However, let's assume that instead of wanting to minimize the total distance traveled to visit each desired destination, we wanted to minimize the total time needed to reach each destination. This goes beyond conventional optimization since travel time is inherently uncertain (traffic jams, time of day, etc.). As a result, to determine our optimal path we would want to use simulation - optimization to first understand the range of potential times it could take to go from one point to another (represented by a probability distribution in this case rather than a specific distance) and then optimize our travel decisions to identify the best path to follow taking that uncertainty into account.

Inverse problems

Probabilistic formulation of inverse problems leads to the definition of a probability distribution in the model space. This probability distribution combines priorinformation with new information obtained by measuring some observable parameters (data). As, in the general case, the theory linking data with model parameters is nonlinear, the posterior probability in the model space may not be easy to describe (it may be multimodal, some moments may not be defined, etc.).

When analyzing an inverse problem, obtaining a maximum likelihood model is usually not sufficient, as we normally also wish to have information on the resolution power of the data. In the general case we may have a large number of model parameters, and an inspection of the marginal probability densities of interest may be impractical, or even useless. But it is possible to pseudo randomly generate a large collection of models according to the posterior probability distribution and to analyze and display the models in such a way that information on the relative likelihoods of model properties is conveyed to the spectator. This can be accomplished by means of an efficient Monte Carlo method, even in cases where no explicit formula for the *a priori* distribution is available.

The best-known importance sampling method, the Metropolis algorithm, can be generalized, and this gives a method that allows analysis of (possibly highly nonlinear) inverse problems with complex *a priori* information and data with an arbitrary noise distribution.

EXERCISES

1. Describe the costs associated with queuing system.

2. Explain the concepts of optimum servicing rate and optimum cost.

3. In a bank operation, the arrival rate is 2 customers/minute. Determine the following:

    i) The average number of arrivals during 5 minutes.

    ii) The probability that no arrivals will occur during the next 30 seconds.

    iii) The probability that at least one arrival will occur during the next 30 seconds.

    iv) The probability that the time between two successive arrivals is atleast 3 minutes.

4. Inventory is withdrawn from a stock of 80 items according to Poisson distribution at the rate of 5 items per day. Determine the following:

    i) The probability that 10 items are withdrawn during the first 2 days.

    ii) The probability that no items are left at the end of 4 days.

    iii) The average number of items withdrawn over a 4-day period.

5. Define a Queue and explain the various queue disciplines.

6. Explain the features and classifications of queuing models.

7. Write a note on various assumptions made in single-channel queuing theory.

8. Explain the terms balking and queue discipline.

9. State the queue parameters.

10. Discuss basic elements of waiting line situations.

11. Explain the Terms: Traffic Intensity, Balking, Reneging and Jockeying.

12. Explain the following with reference to queuing models:

i) M/M/2 ii) Service discipline iii) Kendal's notation.

13. Explain the following:

i) Arrival pattern ii) Service channel

iii) Service distribution iv) Service discipline

14. What is a queuing problem? Explain queuing system, transient and steady state.

15. "The assumptions in queuing theory are no restrictive as to render behaviour prediction of queuing system practically worthless" Discuss.

16. What is the use of single server models?

17. Elucidate the queuing process.

18. What are the different classes of queuing system? Give examples for each of them.