1. BASIC PRINCIPLES

1.1 INTRODUCTION AND PHILOSOPHY

An examination of the origins of any scientific field, be it astronomy or anatomy, physics or psychology, indicates that the discipline began with a collection of observations and results of experiments. It is natural, then, that the first steps in quantifying the subject should involve the collection, presentation, and interpretation of data. Consequently, the study of statistics has played a dominant role in the mathematical preparation of students working in the quantitative areas of the social and life sciences. A statistical treatment of data may be quite elementary, involving little more than listing, sorting, and a few straightforward computations. It may also be quite sophisticated, involving substantial mathematical ideas and delicate problems of experimental design. Once enough data have been collected and adequately analyzed, the researcher tries to identify a process which accounts for the results. It is this activity, the mental or pencil-and-paper creation of a theoretical system that is the topic of this book. In the scientific literature this activity is commonly known as theory construction and analysis. We shall refer to it as the construction, development, and study of mathematical models. As we shall see as our study progresses, the process may involve using familiar ideas in straightforward ways, it may involve using these ideas in unexpected ways, or it may involve generating new concepts and ideas.

The original problem almost always arises in the real world, sometimes in the relatively controlled conditions of a laboratory and sometimes in the much less completely understood environment of everyday life. For example, a psychologist observes certain types of behavior in rats running in a maze, a geneticist notes the results of a hybridization experiment, or an economist records the volume of international trade under a specific tariff policy, and then each conjectures certain reasons for the observations. These conjectures may be based completely on intuition, but more often they are the result of detailed study, experience, and the recognition of similarities between the current situation and other situations which are better understood. This close study of the system, which for the experimenter usually precedes the forming of conjectures, is really the first step in model building. Much of this initial work must be done by a researcher who is familiar with the origin of the problem and the basic biology, psychology, or whatever else is involved.

The next step (after the recognition of the problem and its initial study) is an attempt to make the problem as precise as possible. By this we mean arriving at a clear and definite understanding of the words and concepts to be used. This process typically involves making certain idealizations and approximations. One important aspect of this step is the attempt to identify and select those concepts to be considered as basic in the study. The purpose here is to eliminate unnecessary information and to simplify that which is retained as much as possible. For example, with regard to a psychologist studying rats in a maze, the experimenter may decide that it makes no difference that all the rats are gray or that the maze has 17 compartments. On the other hand, it may be significant that all the rats are siblings or that one portion of the maze is illuminated more brightly than another. This step of identification, approximation, and idealization will be referred to as constructing a real model. This terminology is intended to reflect the fact that the context is still that of real things (animals, apparatus, etc.) but that the situation may no longer be completely realistic. Returning again to the maze, the psychologist may construct a *real model* which contains rats and compartments, but with the restriction that a rat is always in exactly one compartment. This restriction involves the idealization that rats move instantaneously from compartment to compartment and are never half in one compartment and half in another. Also, he might construct a model in such a way that the rat moves from one compartment to another regularly in time, an approximation which may or may not be appropriate depending on just what behavior is to be investigated.

The third step (after study and formation of a real model) is usually much less well defined and frequently involves a high degree of creativity. One looks at the real model and attempts to identify the operative processes at work. The goal is the expression of the entire situation in symbolic terms. Thus the real model becomes a *mathematical model* in which the real quantities and processes are replaced by symbols and mathematical operations. Usually, much of the value of the study hinges on this step because an inappropriate identification between the real world and the mathematical world is unlikely to lead to useful results. It should be emphasized that the construction of a mathematical model is highly non-unique. There may be several mathematical models for the same real situation. In such circumstances, it may happen that one of the models can be shown to be distinctly better than any of the others as a means of accounting for observations. In fact, it often happens that an elaborate experiment is designed for the purpose of showing that one model is truly better than others. Naturally, if this can be shown, then one usually chooses to use the best model. However, it may also happen that each of a number of models proves to be useful in the study - each model contributing to the understanding of some aspects of the situation, but no one model adequately accounting for all facets of the problem under consideration. Thus there may not be a best model, and the one to be used will depend on the precise questions to be studied.

After the problem has been transformed into symbolic terms, the resulting mathematical system is studied using appropriate mathematical ideas and techniques. The results of the mathematical study are theorems, from a mathematical point of view, and predictions, from the empirical point of view. The motivation for the mathematical study is not to produce new mathematics, i.e., new abstract ideas or new theorems, although this may happen, but more importantly to produce new information about the situation being studied. In fact, it is likely that such information can be obtained by using well-known mathematical concepts and techniques. The important contribution of the study may well be the recognition of the relationship between known mathematical results and the situation being studied.

The final step in the model-building process is the comparison of the results predicted on the basis of the mathematical work with the real world. The happiest situation is that everything actually observed is accounted for by the conclusions of the mathematical study and that other predictions are subsequently verified by experiment. Such agreement is not frequently observed, at least not on the first attempt. A much more typical situation would be that the set of conclusions of the mathematical theory contains some which seem to agree and some which seem to disagree with the outcomes of experiments. In such a case one has to examine every step of the process again. Has there been a significant omission in the step from the real world to the real model? Does the mathematical model reflect all the important aspects of the real model, and does it avoid introducing extraneous behavior not observed in the real world? Is the mathematical work free from error? It usually happens that the model-building process proceeds through several iterations, each a refinement of the preceding, until finally an acceptable one is found. Pictorially, we can represent this process as in Fig. 1-1. The solid lines in Fig. 1-1 indicate the process of building, developing, and testing a mathematical model as we have outlined it above. The dashed line is used to indicate an abbreviated version of this process

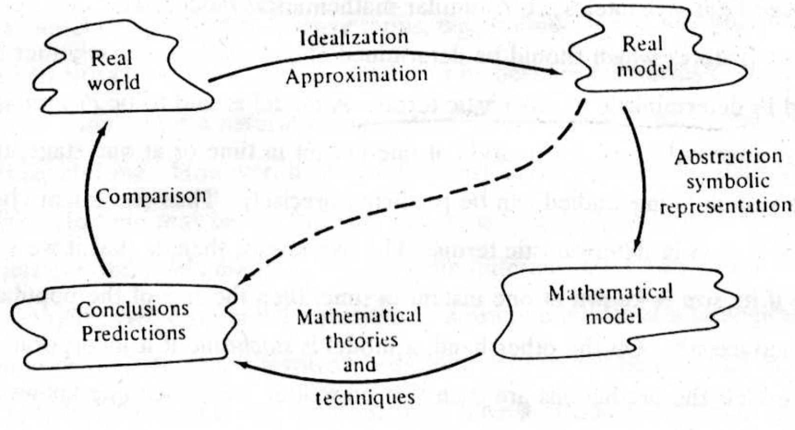

Fig. 1-1

which is often used in practice. The shortened version is particularly common in the social and life sciences where mathematization of the concepts may be difficult. In either case, the steps in this process may be exceedingly complex and there may be complicated interactions between them. However, for the purpose of studying the model-building process, such an oversimplification is quite useful. We also note that this distinction between real models and mathematical models is somewhat artificial. It is a convenient way to represent a basic part of the process, but in many cases it is very difficult to decide where the real model ends and the mathematical model begins. In general, research workers often do not worry about drawing such a distinction. Hence in practice one frequently finds that predictions and conclusions are based on a sort of hybrid model, part real and part mathematical, with no clear distinction between the two. There is, however, some danger in this practice. While it may well be appropriate to work with the real model in some cases and the mathematical model in others, one should always keep in mind the setting that is being used. At best, a failure to distinguish between a real model and a mathematical model is confusing; at worst, it may lead directly to incorrect conclusions. Complications may arise because problems in the social, biological, and behavioral sciences often involve concepts, issues, and conditions which are very difficult to quantify. Hence essential aspects of the problem may be lost in the transition from the real model to the mathematical model. In such cases conclusions based on the mathematical model may well not be conclusions about the real world or the real model. Thus there are circumstances in which it is crucial to distinguish the model to which a conclusion refers.

Many skills are needed for successful model building. In particular the ability to recognize patterns and general structures is just as important in model building as it is in the study of pure mathematics. It is not surprising that the same type of mathematics is involved in studying the servicing of automobiles at a turnpike tollbooth and customers in a barbershop. It is less obvious that, from a mathematical point of view, these two situations have much in common with certain models for the propagation of a rumor or epidemic. However, as we shall see later (Chapter 9), very similar mathematical models will serve for both.

One of the first features which should be determined about a situation is whether it is most appropriately modeled in deterministic or stochastic terms. A model is said to be *deterministic* if it is based on the assumption that given sufficient information at one instant in time or at one stage, then the entire future behavior of the system being studied can be predicted precisely. Thus one might choose to model the growth of a certain population in deterministic terms. The hypothesis, then, is that if we know the way the population grows and if its size is known at one instant in time, then the size of the population for all future times can be determined exactly. On the other hand, a model is *stochastic* if it incorporates probabilistic behavior. For these models the predictions are such that no matter how much one knows about the system at a given time, it is impossible to determine with absolute certainty the nature of the system for future times. For example, suppose that we are concerned with a concept acquisition experiment in which an unbiased subject is provided stimuli intended to convey, in some explicitly defined manner, the concepts roundness, redness, and fourness. Then one might choose to construct a model in which no matter how much we know about how the subjects learned the concepts of roundness and redness does, we cannot determine with absolute precision how quickly they will learn the concept of fourness. The strongest statement that can be made in such situations is a probabilistic one such as, "The subject will learn the concept of fourness with fewer than ten presentations of a stimulus with probability 0.8." Such a model might appear very appropriate in light of the variations shown in experimental results. Many of the most useful models in the social and life sciences are of this type, that is, models whose mathematical description involves chance and uncertainty. Of course, stochastic models have been used successfully in the physical sciences, e.g., statistical mechanics, and deterministic models have been extensively used in the life and social sciences, e.g., Richardson models for arms races and Lotka models for interacting populations. In some instances, one can construct both types of models for the same system, and in some of these cases a comparison serves to check the validity of both (see Chapter 8). The decision as to which type of model should be constructed depends on many factors and is ultimately simply a choice of the investigator. Frequently a deterministic model is taken as a first approximation in a situation when a stochastic model appears more appropriate. For example, when formulated in terms of a real model, a situation involving growth may appear to be best modeled in stochastic terms. However, a stochastic mathematical model may present technical difficulties which are either impossible to overcome or prohibitively time-consuming. It may be desirable to consider a deterministic mathematical model as a first approximation and to compare the conclusions based on such approximations with observations. In general, however, one should not assume that predictions based on one type of model are necessarily better (or worse) than those based on the other type. The relative merits of the two types of models vary from one situation to another.

Some situations lend themselves to descriptions in terms of *continuous* quantities, e.g., space or time, and others are just as naturally phrased in *discrete* terms, e.g., the number of automobiles passing a toll booth in an hour. Even situations which initially appear to be described in terms of a continuous parameter may upon closer examination admit a natural discretization. For instance, a biological population may be viewed as evolving through time. However, if observations are made periodically, then a description of the system in terms of discrete time may be appropriate. There is frequently an option to phrase a situation in one form or the other, and the mathematics is usually quite different in the two cases. For example, difference equations may replace differential equations in a discrete model of a biological system. If computer implemented computations are involved, a discrete version will be required at that step.

If a model is to be of practical use, then one must have a means of obtaining results that can be tested or compared with the real world. Consequently, the model builder must keep in mind the necessity for developing realistic computational schemes or algorithms for computing the quantities arising in the study of the mathematical model. There are some particularly important algorithms associated with the topics discussed in this book, and we shall give an introduction to some of them. Most algorithms of practical significance for real problems require computer assistance with the calculations. Another matter related to checking the model against reality is parameter estimation. Many of the models which are considered here lead to mathematical relations involving a parameter. Tracing this parameter back to the real world, one may find that it is related to a learning rate, a probability of a birth, an asset ratio, etc. Thus, in comparing the results of the mathematical study with reality, it may be necessary to give numerical values to these parameters. In general, the estimation of parameters is a real and delicate problem for which each discipline has its own special and refined techniques. Because of the special nature and limited applicability of many of these ideas, and not because of any lack of importance, we shall consider them only briefly (Chapter 10).

The reader will rapidly realize that some of the best models for certain types of problems are simply intractable mathematically. That is, the model leads to mathematical questions which either have no known solution or no solutions which can be reliably and reasonably computed. In such situations, one often turns to computer simulation for assistance. Again the precise nature of the program depends on the specific problem, and we content ourselves here with some examples (Chapter 10).

To conclude this introductory section, we remark that the use of mathematical techniques, in particular the use of mathematical models, is only one method which can be applied to questions arising in the sciences. As noted earlier, many important aspects of a situation in the social or life sciences may be very difficult to quantify. In such cases the use of mathematical models may be of limited utility, and it may be better to study the situation in the context of a real model by nonmathematical means. Indeed, one might legitimately ask what basis we have for expecting mathematical methods to be effective. Our hopes rest on the proven effectiveness of mathematics in the physical sciences and on scattered but significant successes in the social and life sciences. For example, probabilistic models in genetics (mathematical systems which yield Mendel's Laws as conclusions), logistic models for certain laboratory populations, input-output models in economics (for which Leontief received the 1973 Nobel Prize in economics). However, most of the credibility of mathematical models rests on their unusual effectiveness in the physical sciences and engineering. One of the most impressive examples, indeed perhaps the most impressive example, of the fruitful use of mathematical models occurs in the study of planetary motion. This is also a fine example of the evolution of a model through several stages. Since this example had such a profound influence on science, it is worthwhile to consider it briefly.

1.2 A CLASSIC EXAMPLE

The creation of a coherent system to explain and predict the apparent motions of the planets and stars as viewed from the earth is certainly a significant triumph of human intellect. That the problem had attracted attention from the most ancient times and that the theory is still undergoing modification in this century give an indication of the enormous time and energy that have gone into its study.

Early views of a fixed and flat earth covered by a spherical celestial dome were studied by the Greeks, who devised a real model in the fourth century B.C. which accounted at least approximately for the rough observations then available. The earth was viewed as fixed with a sphere containing the fixed stars rotating about it. The "seven wanderers" (the sun, moon, and five planets) moved in between. The Greeks' concern was to construct combinations of uniform circular motions centered in the earth by which the movements of the seven wanderers among the stars could be represented. Each body was moved by a set of interconnecting rotating spherical shells. This system was adopted by Aristotle, who introduced 55 shells to account for observed motions. This real model based on geometry was capable of reproducing the apparent motions, at least to a degree consistent with the accuracy of the contemporary observations. However, since it kept each planet a fixed distance from the earth, it could not account for the varying brightness of the planets as they moved.

This system was modified by Ptolemy, the last great astronomer at the famous observatory at Alexandria, in the second century A.D. In its simplest form the Ptolemaic system can be described as follows: Each planet moved in a small circle (epicycle) in the period of its actual motion through the sky, while simultaneously the center of this circle moved around the earth on a larger circle. The basic model was capable of repeated modification to account for new observations, and such modifications in fact took place. The result was that by the thirteenth century the model was extremely complicated, 40-60 epicycles for each planet, without commensurate effectiveness.

By the beginning of the sixteenth century there was widespread dissatisfaction with the Ptolemaic system. Difficulties resulting from more numerous and more refined observations forced repeated and increasingly elaborate revision of the epicycles on which the Ptolemaic system was based. As early as the third century B.C. certain Greek philosophers had proposed the idea of a moving earth, and as the difficulties with the Ptolemaic point of view increased, this alternative appeared more and more attractive. Thus, in the first part of the sixteenth century the Polish astronomer Copernicus proposed a heliocentric (sun-centered) theory in which the earth, among the other planets, revolved about the sun. However, he retained the assumption of uniform circular motion - an assumption with a purely philosophical basis - and consequently he was forced to continue the use of epicycles to account for the variation in apparent velocity and brightness of the planets from the earth.

The next step, and a very significant one, was taken by Johannes Kepler. During the years 1576-1596 a Swedish astronomer, Tycho Brahe, had collected masses of observational data on the motion of the planets. Kepler inherited Brahe's records and undertook to modify Copernican theory to fit these observations. He was particularly bothered by the orbit of Mars, whose large eccentricity made it very difficult to fit into circular orbit-epicycle theory. He was eventually led to make a very creative step, a complete break with the circular orbit hypothesis. He posed as a model for the motions of the planets the following three "laws":

1. The planets revolve around the sun in elliptical orbits with the sun at one focus (1609).

2. The radius vector from the sun to the planet sweeps out equal areas in equal times (1609).
3. The squares of the periods of revolution of any two planets are in the same ratio as the cubes of their mean distances to the sun (1619).

These laws are simply statements of observed facts. Nevertheless, they are perceptive and useful formulations of these observations. In addition to discovering these laws, Kepler also attempted to identify a physical mechanism for the motion of the planets. He hypothesized a sort of force emanating from the sun which influenced the planets. This model described very well the accumulated observations and set the stage for the next refinement, due to Isaac Newton.

All models developed up to the middle of the seventeenth century involved geometrical representations with minimal physical interpretation. The fundamental universal law of gravitation provides at once a physical interpretation and a concise and elegant mathematical model for the motion of the planets. Indeed, this law, when combined with the laws of motion, provides a description of the motion of all material particles. The law asserts that every material particle attracts every other material particle with a force which is directly proportional to the product of the masses and inversely proportional to the square of the distance between them. In this framework the motion of a planet could be determined by first considering the system consisting only of the plant and the sun. The latter problem involves only two bodies and is easy to solve. The resulting predictions, the three laws of Kepler, are good first approximations since the sun is the dominant mass in the solar system and the planets are widely separated. However, the law of gravitation asserts that each planet is, in fact, subject to forces due to each of the other planets, and these forces result in perturbations in the predicted elliptical orbits. The mathematical laws proposed by Newton provide such an accurate mathematical model for planetary motion that they led to the discovery of new planets. One could examine the orbit of a specific planet and take into account the influence of all the other known planets on this orbit. If discrepancies were observed between the predictions and observations, then one could infer that these discrepancies were due to another planet, and estimates could be obtained on its size and location. The planets Uranus, Neptune, and Pluto were actually discovered in this manner. However, even this remarkable model does not account for all the observations made of the planets. Early in this century small perturbations in the orbit of Mercury, unexplainable in Newtonian terms, provided some motivation for the development of the theory of relativity. The relativistic modification of Newtonian mechanics apparently accounts for these observations. Nevertheless, one should not view this model as ultimate, but rather as the best available at the present time.

The laws of Newton, viewed as a mathematical model, have provided an extremely effective tool to the physical sciences. The concepts of force, mass velocity, etc., can be made quite precise and the model can be studied from a very abstract point of view. Although the social and life sciences do not yet have their equivalents of Newton's laws, the utility of mathematical models in the physical sciences gives hope that their use may contribute to the development of other sciences as well.

1.3 AXIOM SYSTEMS AND MODELS

In the preceding sections we have surveyed how we intend to use mathematical models to study situations arising outside mathematics, and we considered one example where the use of mathematical models has been particularly rewarding. It is time now for us to begin to make the concept of a model more precise. Since, as we shall see below, a mathematical model can be viewed as an axiom system, it is appropriate to begin with this topic. The following development has been greatly influenced by R.L. Wilder and his book [W].

1.3.1 Axioms

Since the use of the word *axiom* has changed over the years, we begin our discussion of axioms by contrasting the current use of this term with an earlier use. At one time, for example, with Euclid and other Greek mathematicians, the term axiom meant a self-evident truth. It was a universal statement which was obvious to and undebatable by all. An example of such a statement is the proposition "Equals added to equals yield equals." In addition to axioms, mathematicians of the day were also concerned with postulates. These were statements of a more specific character, and they presumably expressed "true facts" about a particular subject, such as geometry. The statement "Through two distinct points there exists one and only one line" qualifies as a postulate. Thus the postulates of geometry use terms such as point and line, which are special to geometry and which are not used in areas such as arithmetic. This separation of basic truths into axioms and postulates has a long history. Euclid used it in his famous Elements, calling the axioms "common notions."

The original use of the words axiom and postulate remained relatively unchanged for almost 2000 years. In fact, it is really only in mathematics itself that a second meaning has arisen. In day-to-day conversation one still hears "It is axiomatic," the meaning being that the statement under discussion is a universal truth. In mathematics, however, a change in the meaning of the term axiom (and likewise of the related term postulate) began during the period which saw the development of non-Euclidean geometries. Without going into detail, we simply point out that a number of geometries were developed in which Euclid's fifth postulate failed to be true. (The fifth postulate essentially said that given a line and a point not on the line, there exists one and only one line through this point and parallel to the given line. Gauss, Lobachevski, and Bolyai all used sets of axioms in which this statement is contradicted while all other Euclidean axioms are true.) The development of these geometries demonstrated the fallacy of the earlier belief that the fifth