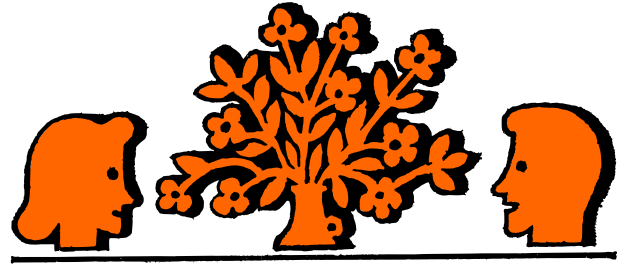
SCIENCE FOR EVERYONE

Ya. I. KHURGIN

YES, NO OR MAYBE



MIR

Science
for Everyone

Я. И. Хургин

ДА, НЕТ или МОЖЕТ БЫТЬ.

Рассказы о статистической теории управления и эксперимента

Издательство «Наука» Москва

Ya. I. Khurgin

Yes, No or Maybe

Translated from the Russian
by Alexander Repyev

Mir
Publishers
Moscow

First published 1985
Revised from the 1983 Russian edition

На английском языке

Contents

Why and for Whom This Book Is Written

An argument once flared up between scientists in a Moscow magazine as to whether or not popular science is necessary and useful. The very statement of the question surprised me a lot. Why then would it never occur to anybody to question the usefulness of travelogues on TV, which acquaint the audience with the life and ways of countries and peoples? But after all the life of the enormous and fascinating world of science is again the life of the world that surrounds us, and the deeper our insight into this world the better. To be sure, a story about a science, especially some esoteric field, is superficially less effective than a story about Bushman tribesmen or the palais and châteaux of France, but the highest achievement of humanity, the interplay of ideas, is no less beautiful. If you are to get a better understanding of the customs and ways of a people, you will first of all be hampered by the language barrier, and so you will have to study the language of the country. Likewise, a "foreign" field of science is Greek to you before you have mastered its language. And if you take the trouble of struggling through the tangle of the language, you will be graced by a pure wellspring of ideas that will now become within your grasp.

A school teacher of mine used to say, "There are no bad words, there are bad mouths." To paraphrase, I would say: "There are no unclear ideas in science, there is lack of desire to make them clear." It is, of course, by no means easy to introduce the nonspecialist reader to the frontiers of modern molecular biology or astrophysics, say. But the chromosome theory of heredity, for example, also at first seemed to be unintelligible and even heretic, but now it is a household word.

I am a mathematician and I have had a happy career having been exposed to radio engineering and physiology, cybernetics and psychiatry, information theory and oil refining, control theory and geophysics. Each of these disciplines, just like a nation, speaks a language of its own, and it takes some time to perceive that, say, radio-frequency pulse and neurone spike, seismogram and control system response are nearly the same notions. Having mastered the fundamentals of different fields of learning, you will eventually perceive that they have much in common, far more than it might appear at first sight.

Major breakthroughs now come from scientific interfaces. Disparate branches of science can mutually enrich one another. That is why I wrote this book, which attempts to give a popular account of an area of science that has experienced two decades of a violent growth. In brief, it can be referred to as the statistical theory of control and experiment.

A popular-science writer contemplating a new book is first of all faced with the problem of selecting material. And so I began by taking down from my shelf a dozen books, thick and

thin, with similar, sometimes even identical, titles and—lo and behold!—I appeared to be confronted with a hodge-podge of topics—so wide is the scope of the science in question. So if I set out to cover all the interesting ideas and techniques, I would have had to give it up as a bad job. I decided, therefore, in selecting and writing to do without references and mostly rely on my own attitudes. I tried to get across to the reader the concepts, having stripped as far as possible my exposition from the fog of terminology and the fat of details. My fight with the terminological excesses, I am afraid, was not always a success, but then every cloud has a silver lining: if this little book generates further interest in the subject and you take it in a more serious way by reading texts and monographs, then such a preliminary introduction to the concepts and terminology will make things easier for you.

Why then was it so difficult to select the material? I hope you have some idea of what control is and of its role in the present-day world. A huge body of literature is devoted to the issues of control without any assumptions about random effects on, or random inputs in, the object under control, be it an aeroplane or a mill, a factor or a state, a living organism or just some abstract object. But how could you possibly describe the control system of an aircraft ignoring atmospheric density inhomogeneities, wind force changes, minor structural inhomogeneities, freight distribution, in-flight passenger motions in cabin, and what not? And how could you possibly describe the control of the vital activity of infusoria

or an elephant ignoring the environmental effects, which vary with time but by no means in a regular way, and all those ups and downs to which our infusoria and elephant have to respond somehow every day, every hour, and every second?

On the other hand, does the above reasoning suggest that generally no control system can be described if we ignore chance? No, it does not, and that is why.

We are all accustomed to the variation of day and night length: from 22 June to 22 December the day becomes shorter, and then the variation reverses. And this all is on a strictly regular basis, accurately predictable for years to come. Just imagine our life with random day length and wanton changes of day into night, that is, if the rotational velocity of the Earth spinning on its axis were not constant, but changed arbitrarily like the mood of a nymph. Now the sun quickly rose and you hurry to work, but the day drags on and on—the Earth's rotation slackened suddenly. Now the rotation hastened unexpectedly, and you broke your date made for just before the sunset. The evening flew by, but you did not have enough sleep—the night was only three hours long. Next came a short day, and no sooner you had your lunch than the night fell, this time a long one... All the reflexes of sleeping and waking hours are in turmoil. A wretched life!

As a matter of fact, the Earth spins about its axis in a slightly irregular manner: now a meteorite hits it, now a comet flies by—but these impacts are negligible and so their influence upon the day length and alternation of day and night

is negligible too. To all intents and purposes, in our everyday life we can think of the Earth's rotation as absolutely regular.

Air temperature is subject to far more noticeable variations. But we have learned to cope with them, by continually controlling the behaviour of ourselves, our children and subordinates. So we put on warm things, use an umbrella, open or close a window, turn on the central heating or mend the roof. Our life would be much easier if the air temperature on Earth varied in a regular way, e.g. according to a sine law, falling off from $+25°C$ at summer solstice to $-25°C$ at winter solstice and back. No unexpected cold spells, no problems with attire—when to buy a fur coat or a bathing suit, and so forth. Accordingly, the question of whether or not we are to take into account random inputs in any problem, including control problems, should be approached with caution: in some cases we can very well ignore random inputs, in others not. But it so happens that situations coming under the last heading are legion, and so we chose them as a subject of this little book.

It is only natural to describe random inputs and perturbations drawing on the results of a science concerned with random events, quantities and processes, that is, probability and mathematical statistics.

The book would be of especial value for students with some knowledge of probability. Nowadays probability is included not only in a college mathematics course but also in the curriculum of high schools in many countries. And so, I hope, this book will appeal to a wide range of readers.

But still we should make allowances for the fact that, taking some material without any practical aim in the offing to which this material might be applied usefully, students generally tip-toe their precious knowledge only as far as the examiner's desk, and, having got the tiptop mark and heaved a sigh of relief, they almost instantly relieve their memory of unnecessary burden. With this in mind, the book provides a quick introduction to the elements of probability.

Frankly speaking, there was another incentive for writing this book. In 1974 my book *Did You Say Mathematics?* was issued by Mir Publishers. This was a book of essays about mathematics in general, its methods and ideas, and the relations of mathematics with other sciences. Judging from the responses of the readers, it appeared of help for biologists, engineers, economists, chemists.

The present book is somewhat different in character. It is devoted to a definite mathematical discipline. Since probability and statistics occupy a fairly large place in our life, I would be happy if this book would be useful to many lay and specialist readers.

The Author

Uncertainty and Randomness

Most of the current books for the nonspecialist reader using some notions of probability and mathematical statistics normally begin by introducing the elements of the theory of probability such as probability, conditional probability, probability distribution, random variable, mathematical expectation, variance, and so on. I do not want to move in this rut because, as my teaching experience shows, it is the exposition of the ABC of probability in several pages that usually produces in the reader the illusion that he (or she) has already mastered the essentials, whereas precisely the initial premises and fundamentals of probability and mathematical statistics, the issues of applicability of the theory and paradoxes, meet with significant psychological difficulties, especially in people long past their student age. At the same time, the formalism, i.e. mathematical machinery, of the theory of probability is essentially similar to calculus and linear algebra, and presents no difficulties. Concise introductory guides, as a rule, only contain a short section devoted to the initial

concepts underlying the theory of probability. This book, by contrast, takes care to expose them in considerable detail.

... You go out and run into a blonde. No, not the blonde who made you open your eyes the other day, but just a blonde, that is, not a redhead, or a brunette. In the parlance of probability your meeting the blonde is an event, which may or may not occur, and that is it. But in terms of everyday life, your meeting the blonde may be quite an occasion, or may even be unpleasant, or else may be of no significance—we will be looking at the importance of events in a section on risk.

If each time you leave your home you record whether or not you first meet the blonde, you will be able to calculate the frequency of the event the first person you meet is the blonde (the frequency is the ratio of the number of times you meet the blonde to the total number of observations). Above all, notice that you are able to make your observations repeatedly under identical conditions. Summer or winter, sunny day or rainy evening, the chance of the event is generally the same.

It is here assumed that you are able to make your observation indefinitely. Quite likely, as the total number of observations increases, the frequency of the event will vary but little and nonsystematically, and if the monthly number of observations is the same, the frequency again will fluctuate only slightly. Let the number of observations be large and you preselect some subset of observations, e.g. each third or first one hundred fifteen in each thousand, such that the number of

observations in it is sufficiently large and grows infinitely with total observations. Then the frequencies derived both from the subset and the total number of observations will be similar. If an event displays such properties, it is called *random.* It is under these conditions that the concept of probability of the occurrence of the event is introduced as an axiom, just as the limit of the frequency of its occurrence.

It is impossible to predict the outcome of gymnastics events at Olympic Games, as much is dependent on a chance. So we once witnessed uneven bars collapsing during the performance of a favourite. Injuries and illnesses are possible, some performers may be not at the peak of their form... And still, in terms of the theory of probability, the Olympic scoring is not a random event: you cannot repeat an Olympiad indefinitely under identical conditions. Next games will have other participants, they will be conducted at another site, and so forth. Such events are termed uncertain, rather than random. Mathematically, they are taken care of by game theory. But this book will concentrate on random events, probability and mathematical statistics, which are concerned with them.

To sum up, not every event whose result is unknown and does not lend itself to unambiguous prediction may be called random in the language of this book. An event becomes random under certain conditions and constraints just described. I will refer to this property as *statistical stability.* So the book discusses the theory of random, statistically stable events.

Control

You are likely to have read that enormously popular novel by Jules Verne entitled *Les enfants du capitaine Grant*, but I want here to recapitulate the story.

In the summer of 1864 the beautiful yacht *Duncan*, at the end of its test voyage, was sailing past the island of Arran.

The proprietor of the yacht Lord Edward Glenarvan, one of the sixteen Scottish peers who seat at the House of Lords, was completing his travels together with his young and charming wife Helena, cousin major MacNabbs and captain John Mangles.

A watchman noticed astern a huge balance fish. The fish was caught, gutted and a strong bottle was found in it. The bottle was broken and there some pieces of paper were found, which were badly damaged by sea water.

After having examined the scraps Glenarvan said: "Here are three documents, obviously copies of the same text. One is written in English, another in French, and yet another in German". Now the whole team set out to recover the text, comparing the three versions, and in a time they produced the following enigmatic text:

On 7 June 1862 the three-mast ship Britannia *Glasgow*
was wrecked gon austr
shore two sailors
Capitain Gr rea
conti pr cruel Indi

thrown this document longitude
and 37°11′ latitude Render them assistance
perish

It was then necessary to decipher the note, thinking of the absent parts of the text. Those who are fond of cross-words know the problem.

"After a moment's silence Glenarvan went on to say:

'My friends, these all suppositions seem to be quite reasonable. I think the disaster took place off the Patagonian shores.'"

They found from the newspapers:

"On 30 May 1862. Peru, Callao. Destination Glasgow, *Britannia*, captain Grant."

"'Grant!' exclaimed Glenarvan, 'If it isn't that galant Scotchman who day-dreamed of founding a new Scotland on one of the islands in the Pacific?'

'Quite so,' said John Mangles, 'the very Grant. In 1861 he got underway from Glasgow on *Britannia*, and since then there has been no sign of him.'

'No doubt,' cried out Glenarvan, 'it's him! *Britannia* left Callao on 30 May, and on 7 June, in a week's time, she was wrecked off the Patagonian shores. We now know the whole story of the disaster. My friends, you see we have found a clue almost to all the puzzle, and the only unknown here is the longitude of the place of the wreck.'

'We need no longitude,' said captain John Mangles. 'Knowing the country and latitude, I undertake to find the place.'

'And so we know everything?' asked Lady Glenarvan.

'Everything, dear Helena, and I can fill in the gaps produced by sea water with such an ease as if the document were dictated by captain Grant himself.'

And here Glenarvan took the pen and without hesitation wrote the following:

'*On 7 June 1862 the three-mast ship* Britannia *of Glasgow, sunk off the shores of Patagonia in the Southern hemisphere. Two sailors and captain Grant will try and reach the shore where they become prisoners of cruel Indians. They threw this document at longitude and 37°11' latitude. Render them assistance, or they will perish.*'

'Well, well, dear Edward!' cried out Lady Helena, 'If those wretches are to see their native shores again, they'll owe you their salvation' ".

Thus Glenarvan put forward a hypothesis as to the place where the *Britannia* was wrecked, and after he had met the children of captain Grant he organized a voyage to Patagonia. The purpose of the voyage was the search for the lost expedition, or, if we are to use the dry language of science, the test of the hypothesis, and should it come true, the rescue of the ship's crew. Let us now skip about 200 pages of breath-taking adventure and review the situation. Grant's team was not found in Patagonia, and so Glenarvan's hypothesis turned out to be wrong. It is here

that a chance member of the crew, the famous geographer Paganel, suggested another interpretation of some of the word scraps in captain Grant's note:

"Passing his finger over the scrappy lines of the document and underscoring some of the words confidently, Paganel read the following:

'*On 7 June 1862 the three-mast ship* Britannia *of Glasgow was wrecked after...* (Here, if you wish, you may insert *two days*, *three days*, or *long agony*—all the same.)... *off Australian shores. Heading for the shore*, *two sailors and captain Grant tried to land...*, or *landed on the continent*, *where they became prisoners of cruel natives. They threw this document...*' and so on and so forth."

Another hypotheses was put forward, and the leader of the expedition takes another decision—the *Duncan* makes her way to Australia. Again we will skip 400 pages of exciting and riveting adventure. But in Australia captain Grant had not been found as well, and so Paganel's hypothesis proved wrong.

The former boatswain of *Britannia* Ayrton supplied some new piece of information: shortly before the wreck captain Grant planned to visit New Zealand. But even before Ayrton's report came Paganel understood that his interpretation had been erroneous and suggested a fresh version:

"'*On 7 June 1862 the three-mast ship* Britannia *of Glasgow*, *after a long agony was wrecked in South-*

ern Seas, off New Zealand. Two sailors and captain Grant managed to reach the shore. Here suffering cruel deprivations they threw this document at... longitude and 37°11′ latitude. Render them assistance, or they will perish.'

There was a silence. Such an interpretation of the document was again possible. But for exactly the same reason that it was so convincing as the earlier interpretations, it could be also erroneous."

There is hardly denying that any hypothesis, even quite a reasonable one, might turn out to be false under test. In fact, the last rendering of the text, as suggested by Paganel, appeared to be false due to the falseness of the very first hypothesis of Glenarvan. Remember that he supposed that the texts written in the three languages were absolutely identical, which was not so, since the island where Grant landed was the island of Maria-Theresa on English and German charts, the Tabor on French charts.

Let us summarize the behaviour of Lord Glenarvan. He comes in possession of a piece of evidence, analyzes it, suggests a hypothesis and makes a decision concerning actions to test it. As a result, he obtains new information used as a foundation for accepting or rejecting the hypothesis, since it may appear true or false. Next, depending on the results of the test and fresh information, Glenarvan reconsiders the situation, puts forward fresh hypotheses and makes a decision, and so forth. The procedure is followed till the aim is achieved or it is found

that it is impossible to do so, and thus a decision is taken to discontinue the search.

This is precisely what is meant in science by control. Control of a ship, factory, oil refinery, school, and any other object occurs in a like manner. At first hypotheses are suggested (a ship either is steady on course or goes off it), next the hypotheses are tested based on the information available (observations are made, parameters are measured), and lastly decisions are taken as to measures required.

The coronation of a British queen, a church sermon, or an execution of the opera *Eugene Onegin* are all examples of events without any uncertainty: all the actions and words are predetermined, and there is, in essence, no control. Admittedly, when in the last act of the opera the mourning Tatyana sinks into an arm-chair to enable Eugene to knee before her, you might think of a possibility for a pussy-cat to have curled cosily in the arm-chair with all the disastrous implications. But still, despite the uproar in the house, Tatyana would not smile and, after the cat had gone, would continue to suffer and Onegin would still sing his final aria.

Such situations go to prove that control presupposes some uncertainty and a possibility of choice.

Therefore, speaking about control, we will tacitly imply the presence of uncertainty in which sequentially hypotheses are made and tested, decisions are made and tested. In what follows we will take a closer look at certain aspects of this fairly complex process.

Henry Adams Takes a Decision

Let us recall the beautiful Mark Twain's story *The 1,000,000 Bank-Note.*

A mining-broker's clerk in San Francisco sailing on a little boat on the bay, ventured too far, was carried out to sea, and was picked up by a small brig bound for London. Adams had to work his passage as a common sailor. When he stepped ashore in London his clothes were ragged and shabby, and he had only a dollar in his pocket. Next day, hungry as a wolf, he was fiddling about near a manor-house, where the following events had been taking place:

"Now, something had been happening there a little before, which I did not know anything about until a good many days afterward, but I will tell you about it now. Those two old brothers had been having a pretty hot argument a couple of days before, and had ended by agreeing to decide it by a bet, which is the English way of settling everything.

"You will remember that the Bank of England once issued two notes of a million pounds each, to be used for a special purpose connected with some public transaction with a foreign country. For some reason or other only one of these had been used and cancelled; the other still lay in the vaults of the Bank. Well, the brothers, chatting along, happened to get to wondering what might be the fate of a perfectly honest and intelligent stranger who should be turned adrift in London without a friend, and with no money but that million-pound bank-note, and no way to account for his being in possession of it. Brother A said

he would starve to death; Brother B said he wouldn't. Brother A said he couldn't offer it at a bank or anywhere else, because he would be arrested on the spot. So they went on disputing till Brother B said he would bet twenty thousand pounds that the man would live thirty days, *anyway*, on that million, and keep out of jail, too. Brother A took him up. Brother B went down to the bank and bought that note. Then he dictated a letter, which one of his clerks wrote out in a beautiful round hand, and then the two brothers sat at the window a whole day watching for the right man to give it to."

And so Henry Adams happened to be that stranger. He was interviewed, given an envelope and said he would find the explanation inside, he should take it to his lodgings, look it over carefully, and not be hasty or rash. The hero goes on to tell:

"As soon as I was out of sight of that house I opened my envelope, and saw that it contained money! My opinion of those people changed, I can tell you! I lost not a moment, but shoved note and money into my vest pocket, and broke for the nearest cheap eating-house. Well, how I did eat! When at last I couldn't hold any more, I took out my money and unfolded it, took one glimpse and nearly fainted. Five millions of dollars! Why, it made my head swim."

So two absolutely opposite versions of his fate loomed in his dimmed mind. And he could make, as the two gentlemen had made, two hypotheses about his life during the period he was in possession of the bank-note: a failure, when he would be required to provide explanation about

where he, a tramp, got the note; a success, if he manages to make use of this money.

Hypotheses are denoted by the letter H. The initial hypothesis, or null hypothesis, "Henry Adams will fail" will be denoted by H_0, and the competing, or alternative, hypothesis, opposite to the initial one, "Henry Adams will be a success" will be denoted by H_1.

Henry's situation was desperate: he always had to make a decision, which might bring him either to a failure or a success.

And so Henry wanders about streets. He sees a tailor-shop and feels a sharp longing to shed his rags and clothe himself decently once more. Now he has to make a decision. If he says "yes", i.e. expects being exposed in visiting the shop, and accepts hypothesis H_0, then he does not enter the shop and keeps on his rags. If the says "no", i.e. accepts hypothesis H_1, he enters the shop and asks to sell him a ready-made suit.

Whatever his decision, Henry may make a blunder. What errors are possible? One at which the null hypothesis is true (Henry Adams is bound to succeed), and accepted is alternative hypothesis H_1, is called the first-type error, or the error of the first kind. He believes in his lucky star, and so he enters the shop. But if, when the tramp hands the £ 1,000,000 bank-note to a shop-assistant, the latter will take him for a thief, Henry will make an error of the first kind. But if alternative hypothesis H_1 is true, but Henry accepts H_0, he will make an error of the second kind.

In this situation, if Henry shrinked from entering the shop for fear of being mistaken for

a thief (but if he did approach the tailors and produced his large note, and they, stunned by the appearance of a millionaire, would rush to provide him with garments on credit), then Henry's misgivings would turn out to be false. This would exactly be the second-type error, or false alarm.

The keen reader may say that we could think of Henry Adams's success as the null hypothesis, and hence his failure as the alternative one. Quite so. The error of the first kind is often taken to be the one that is more important to avoid, although this is not always the case. If the errors are of about the same significance, which is rarely so, then it is immaterial which is taken to be which. Let us summarize the versions possible in the table.

Table 1

Henry's decision	Reality	
	Failure	Success
Failure	True	Second-type error — false alarm
Success	First-type error — omission	True

Although for Henry Adams errors of the first and second kind are by no means equivalent, he, being a hazardous person, selects hypothesis "success". He comes through with flying colours,

because the personages of the story are stunned by the looks of the million-pounder and their respect for it is unbounded.

This exciting story may serve as an illustration of a situation often encountered in everyday life: on the basis of some consideration or evidence several hypotheses are made, and so an observer or a manager, a research worker or a gambler has to decide on one of those hypotheses and shape his plans accordingly.

Recall the statistically stable event—to meet a blonde first when leaving your place. Such hypotheses may be made: the probability of the blonde turning up first is less than 0.1, since black-haired girls occur more frequently; the probability of encountering three blondes one after another is more than 0.01. Such hypotheses are called statistical, since they involve statistically stable events, and lend themselves to a test by statistical means.

At the same time the hypotheses that Henry Adams will not starve to death having a million-pounder in his vest-pocket or will not end up in a jail, but will survive and be a success, are not statistical hypotheses, and so they cannot be tested statistically, because the event "Henry Adams is a millionaire" or "Henry Adams will fail" is an uncertain event, and not a random one. And in making his decisions Henry relies on intuition, not statistical data. In later sections we will be looking at situations where hypothesis testing and decision making rely on observational and experimental evidence, not intuition and where mathematical statistics comes in.

A Glimpse of Criteria

When in the summer of 1973 Hammer's private collection of pictures was brought to Odessa in the Ukraine, I was so fascinated by the exhibition that visited it several times. In addition to the pleasures derived from viewing the famous masterpieces I was graced by a variety of biting, acid and witty comments of the Odessa public, who in the Soviet Union have a reputation for wit and temperament. Opinions differed widely, and if one was delighted with Van Gogh's *Sowers*, another criticized severely the violet colouring of the picture and stated that he would have never hung it in his home, not for the world.

Opinions thus appeared to be highly subjective, and it was absolutely impossible to work out any criteria: some mentioned the sluggishness of drawing, others the masterly saturation of colours, still others the oversaturation—so many men, so many minds.

Suppose 13-year old pupils of two classes A and B start an argument as to which class is taller. It is hard to decide.

The A pupils cry that their John is taller than any of the B pupils. But the latter counter that this proves absolutely nothing. Just one giraffe in a class, what of it... But all the other B pupils are taller. Such an argument customarily results in a brawl, mainly because the subject of the difference is completely obscure.

In fact, what is to be understood under the height of a class? What meaning do the children attach to this quantity? How is it to be defined? Since normally nobody comes up with consistent

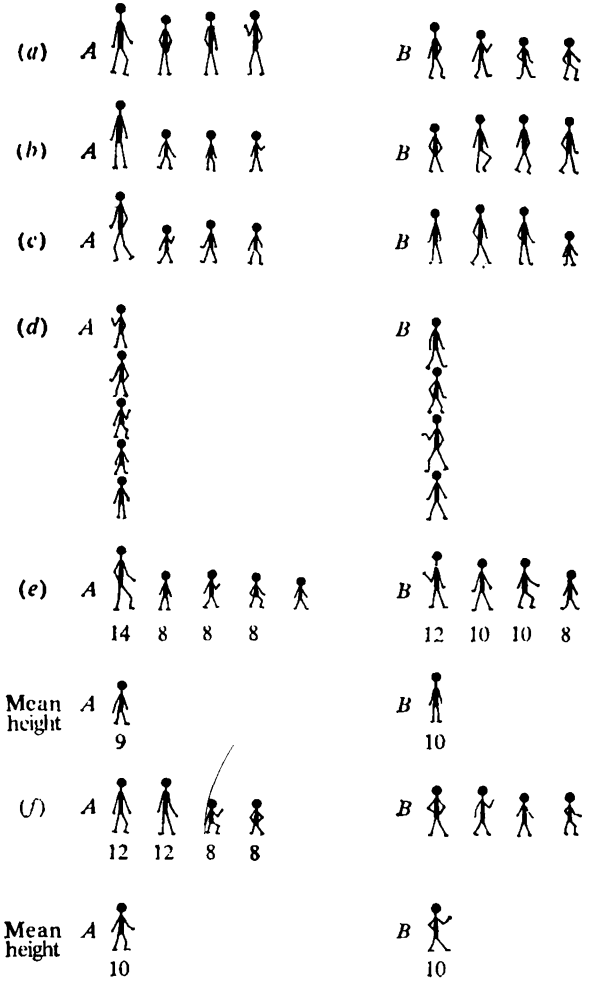

(a) A
B
(b) A
B
(c) A
B
(d) A
B
(e) A
14 8 8 8
B
12 10 10 8
Mean height A
9
B
10
(f) A
12 12 8 8
B
Mean height A
10
B
10

Fig. 1

answers to these questions, I would suggest some options. These are as follows:

A pupils, or A's, are considered taller than B's, if:

(a) *Any of the A's is taller than B's* (Fig. 1*a*). The situation is self-explanatory, but unlikely.

(b) *The tallest of the A's is taller than the tallest of the B's* (Fig. 1*b*). It may so happen here that one of the A's is the tallest, but all the remaining ones are short, e.g. all the other A's are shorter than any of the B's. This option of the answer should, I think, be rejected, although in a basketball match between classes it is the tallest who may dominate.

(c) *For any of the A's there is a shorter B pupil* (Fig. 1*c*). In that case, among the B's there may be one squab, and although the other B's are taller, even taller than all the A's, except for the big one, the A's are still ahead. Intuitively this option also does not appeal to our understanding of the height of groups of people.

(d) *The sum total of heights of the A's is larger than the sum total of heights of the B's.* In Fig. 1*d* the A pupils appeared to be taller. Such a situation may come about for a great variety of reasons, e.g. since each of the A's is taller, but primarily the situation may be caused by the fact that there are more boys in A than in B. Perhaps in this situation such a criterion may not appear to be relevant, but in the case of tug of war, say, if we were to compare the strengths of the classes, this criterion—the sum total of strengths—would be quite satisfactory.

(e) *The mean height of A is larger than the mean height of B.* We will thus have to work out

the arithmetic mean of heights of each of the classes and then compare the results (Fig. 1*e*). Now everything boils down to the spread of heights in a class, and the A pupils lose. The giraffe does not settle the matter, because the other A's are shorter by far, whereas the B's are a bit higher than most of the A's.

(f) The dispute, you see, is really difficult to work out: arguments are also required for selecting one of the criteria. The situation becomes especially involved in the case of the situation in Fig. 1*f*: the mean height of both classes is the same, but the B's are of about similar height, whereas the A's include both tall and short pupils, who compensate for one another.

If we again review all the options, the mean height option will still appear to have more intuitive appeal, and so it would be better to assume this option as a criterion.

The height of pupils here should, of course, be treated as a random variable, and therefore the mean height is a sample mean, or empirical average, which represents the *mathematical expectation* (or just *expectation*) of the random variable itself, i.e. of the height of the pupils. We agreed from the very beginning that the concept of expectation is considered known, and so the above is just some generalization.

Among the pupils there are fair, red, brown, and black-haired persons, and so we can consider the *probability distribution* of another random variable, the height of red-haired pupils. Now this will be a *conditional distribution*: the probability distribution of the height, given that the pupils are red-haired. Generally speaking, the condi-

tional distribution and the initial, unconditional distribution will be different. Each of them will have its own expectation, since the mean heights of the 13-year old pupils and red-haired 13 year-olds may be unequal.

The mean height of red-haired pupils here is the *conditional expectation*—another criterion, say, for comparison of the height of red-haired 13-year olds from different schools. The conditional expectation is an important concept and we will make much use of it throughout.

When a teacher in a class gives marks for a dictation, he has to sum up the total number of mistakes and give A if there are no mistakes, and D if there are many mistakes—on the face of it everything is okay here: the criterion is clearly defined. But the teacher, consciously or subconsciously, distinguishes between blunders and slips, mis-spellings and omissions, and at times the progress or setback of a pupil, so "violating" the rules. I am not going to blame the teacher for such a practice. My more than thirty five years of teaching experience show that the essentially informal instructive activities are poorly subject to any formalization, and all sorts of red-tape instructions here are hindrance, rather than help, for a skilled instructor, and young teachers do not need instructions of the traffic-rules type, they rather need friendly guidance to improve their teaching qualification and practices.

Qualification of knowledge of students using the four-level marking system is thus a test of the hypothesis that a given dictation or a given student can be referred to one of those grades. No matter how inadequate the mark criterion is,

it is still there and can be expressed by a number.

Sometimes, however, two marks are given—one for spelling, the other for punctuation—or even three, e.g. for an essay, when the story is also evaluated. What criterion would you suggest to compare essays and select the best one? How, for example, would you go about selecting the best one among the three essays with the following (1) number of spelling mistakes, (2) number of punctuation mistakes, and (3) story mark (where 5 means excellent, 4 good, 3 fair and 2 bad): 2/1/4, 0/1/3, and 4/0/5. Quite a problem, isn't it?

Precisely because the possible solutions are so ambiguous, the criterion must be expressed by one number.

Let us try and think of some criterion that would enable us to solve the problem of selecting the best essay.

Note that the larger the first two numbers (spelling and punctuation mistakes) the poorer the essay, and the higher the story mark the better. We can here take as a criterion the ratio

$$R = \frac{\text{Style}}{\text{Sp}+\text{P}+1},$$

where the unit in the denominator is added so that R would not become infinite when there are no mistakes at all.

In our three cases we will thus have

$$R_1 = \frac{4}{2+1+1} = 1, \quad R_2 = \frac{3}{0+1+1} = \frac{3}{2},$$

$$R_3 = \frac{5}{4+0+1} = 1.$$

Consequently, according to criterion R the second work is the best, the first and third being equal.

But I somehow appear to favour the last work: the top mark for the story—an appreciable improvement over the fair mark, even if there are a bit more mistakes. And so another criterion suggests itself, which favours the story mark: instead of one in the denominator we will have five. This gives

$$K = \frac{\text{Style}}{\text{Sp}+\text{P}+5}$$

and

$$K_1 = \frac{4}{2+1+5} = \frac{1}{2}, \quad K_2 = \frac{3}{0+1+5} = \frac{1}{2},$$

$$K_3 = \frac{5}{4+0+5} = \frac{5}{9},$$

i.e. according to criterion K the last work takes the cake, the first and second being equal.

Now, I think, you can easily suggest criteria according to which the first place will go to the first work, or else all will be equivalent. So you can get anything that suits you.

To summarize, in a problem on selecting the best object or best solution it is always possible to suggest a wide variety of criteria. You have thus seen how critically the fate of an object at hand, a dispute, or a competition of works of art, even a human or a group of people is influenced by the criterion selected, and how ephemeral are at times the speculations about fair and unfair solutions, when both the situation is evaluated and criterion is selected in a fairly arbitrary manner.

Radar

When in pitch darkness you stump through the construction site and the battery of your flash-light has run down, you turn on the light intermittently, for a short time—just to be able to see if the path is free or not, in order not to bump into something or break your leg. If there is something ahead, the ray from the flash-light will make a light spot and so you will see an obstacle, although you will not be able to say with certainty whether it is a tree, a concrete slab or a car. If there is no obstacle, the ray will disappear into the darkness and you can safely make a few steps ahead. But you may make a mistake: stop short in fear when in fact it will be a reflection from a distant window or an inclined piece of roofing iron, the reflected light spot will strike sideways and you will, cursing, hurt yourself at something.

So we have two hypotheses: H_0—no obstacle, and H_1—there is an obstacle. We can make mistakes of the first and second kind: stop in fear of an obstacle, when there is actually none, and overlook an obstacle, when there is actually one.

Let us now look at a more serious problem, that of taking a decision in radar. The primary task of radar is to detect aircraft when they appear within an acquisition range. To be sure, radar also performs other functions: it determines the coordinates and velocity of a target, it can do other things depending on the purpose of the radar station. But we will concentrate for the moment on the problem of detection alone. By

the way, ship-borne radars must also detect other ships, icebergs and outline shorelines.

Radar consists essentially in sending out pulses of high-frequency radio waves, and receiving some of the radiation reflected by the target. The receiving aerial is the radio counterpart of the eye. The received signal is very weak and so it needs high amplification. But the problem is complicated by internal, or receiver, noise, atmospheric noise, and other outside interferences. These all may lead to errors.

Before we turn to these errors, let us consider the detection problem in air defence, which is more crucial than in air traffic control. From all over the area air defence system the alert defence centre continually receives information about the presence of all the target (aircraft and missiles) within the area.

Even the most advanced and expensive systems are not error-free: target signals may be mistaken for noise, which is always there and is the background against which the signal must be detected. This is the error of the second kind—the omission of vital information, especially vital in the days of supersonic velocities.

A false alarm, an erroneous decision that a hostile object has been detected when there is none in actual fact, is not as innocent as it might appear: this may cause the air defence force take retaliation measures, which is no good at all.

A radar system can be organized differently not only in the context of hardware, but also in the context of data processing techniques, and thus the characteristics of different systems will

be different as well. How are we to judge about the quality of the functioning of a radar system?

In each cycle a transmitted pulse may reflect from an object and be either received or omitted; if there is no target, the pulse will go away and the receiver may either indicate no-signal or mistake noise for a signal. It is worth reminding here that noise is a random process, and the radar operates under the conditions of statistical stability. Therefore, hypothesis H_0 (only noise) and hypothesis H_1 (noise and signal) are statistical hypotheses and it here makes sense to speak about the probabilities of errors of the first and second kind. In this case the error of the first kind occurs when there is no object, or rather there is only noise, and H_1 is accepted, i.e. false alarm. In mathematical statistics the probability of first-type error, or the probability of false alarm, is called the *significance level.* This term is more apt, since the significance of making a false decision about the presence of an object, where there is none, is quite clear, and here it is only improved by being quantified.

We can now represent the situation in the form of Table 2, similar to Table 1.

Notice, however, a substantial difference between the two tables: although Henry Adams can estimate his loss for both errors, there are no probabilities of these errors, since the event "Henry is put to jail" is an uncertain event, not random, and Henry's behaviour, although he performs in an arbitrary manner, have no probability distribution.

The radar problem thus boils down to favouring one of the two hypotheses, and making

Table 2

Decision	Reality	
	Noise	Signal+noise
Noise	True. True decision probability $1-\alpha$	Second-type error—signal omission. Probability β
Signal+noise	First-type error—false alarm Error probability—significance level α	True. True decision probability $1-\beta$

one of the two possible decisions, to say yes or no.

A wide variety of ways to make a decision in this situation are possible. We can, for example, favour a method that provides the lowest significance level, i.e. the lower first-type error.

Before discussing the possible values of the significance level, we will sketch other problems associated with statistical tests of hypotheses.

You will have heard about the Morse code, in which letters and numerals are represented by dots and dashes. We could replace them by any other two different signals. In modern telegraphy they use either mark pulses and pauses or d.c. mark pulses of different polarity or a.c. pulses of the same length but different frequency and phase. The main thing about all methods of coding is the use of two distinguishable signals.

Using computer language, we can represent them as 0 and 1. Each letter or numeral will then be a combination of characters, 0's and 1's.

If characters are transmitted over a communication channel, then in any of the above-mentioned techniques, owing to the presence of interferences, a 0 may be deciphered as a 1, or vice versa. Formally, the situation is as in the radar problem: H_0 is the transmission of a 0 and H_1 is the transmission of a 1, and the first-type error is to mistake a 0 for a 1, and the second-type error is to mistake a 1 for a 0. The significance level here is the probability of receiving a 1, when a 0 was transmitted.

A similar situation occurs in quality control. The figures of merit here vary widely. So for shoes, tyres, incandescent lamps the main criterion is the service life.

The paradox of the situation is that, for one thing, you do not want to buy shoes that will only live for a week or a month, for another, you cannot test the shoes for durability without wearing them out—the proof of the pudding is in the eating.

The life of products is generally determined by analogy. From a large batch of products we take a sample, i.e. select some tyres following some rule, and test them for durability. Tyres, for example, are installed on control cars to test what distance covered will make them bald. If almost all the tyres of the sample made, say, one hundred thousand kilometres, it is quite probable that all the other tyres in the batch will perform in the same way. In other words, we assume that the products of the batch all have

the same properties (as far as their quality is concerned) as the products of the sample.

In quality control there are two hypotheses: the batch is good (H_0), or the batch is bad (H_1). Control consists in testing the hypotheses and, of course, taking one of the two possible decisions: go or no-go, yes or no. A complete analogy with the radar problem.

To summarize, any quality control method is testing statistical hypotheses to make decisions under uncertainty conditions, which at that are characterized by statistical stability. This way of looking at things leads to a far-reaching interpretation of mathematical statistics. In the literature of last decades you may even come across such a definition: the subject of mathematical statistics is finding rules of decision making under uncertainty characterized by statistical stability. This alone would, of course, be sufficient to class mathematical statistics among the key disciplines.

To return to quality control, the figures of merit are predetermined by the purpose of products. So in matches or nails the acceptable reject level may be 5 per cent, but in aircraft engines or medicines the quality standards are higher by far. Hence the difference in requirements to quality control methods.

As it was said above, any of alternative hypotheses can be taken to be the null hypothesis, the decision being made by the analyser himself.

We will now approach the situation from the side of the supplier for whom a rejection of a good batch is the most unfavourable event, and so the rejection of a good batch will here be

assumed to be the first-type error. Let the probability of first-type error be α, then in a given set of tests the percentage of rejected true hypotheses will be 100α. So, at $\alpha = 0.02$, $100 \times 0.02 = 2$, and true hypothesis rejection will average 2 per cent. A measure of confidence that H_0 is true (the batch is good) is here the probability $1 - \alpha$, called the *confidence level.*

Likelihood Ratio

You will have heard of acceleration in humans (from the Latin word *acceleratio*), which means a quickened development of the human body in, as anthropologists put it, certain ethnical and professional groups of population and earlier puberty.

Long-term observations indicate that the distribution of the height of people does not vary from generation to generation. This distribution is normal. It is natural, therefore, to see if the parameters of the distribution change or not. Let us only consider the mean, i.e. find out if the mean height changes from generation to generation assuming the normal height distribution.

On the fairly reliable statistical evidence concerning the age groups born in the years 1908-1913, we can find the mean height of adult men in the age from 20 to 30, it is 162 centimetres. For the years 1943-1948 we obtain 170 centimetres. We are now to test the hypothesis that the mean height does not change from generation to generation, and the fluctuations observed are the "devilish plot", i.e. just a natural spread, since

the mean height is a random variable. Our null hypothesis will be that the mean height in the 1943-1948 age group is the same as in the 1908-1913 age group, the alternative hypothesis being that the mean height has increased over these 35 years.

What follows is an illustration of the reasoning used in such situations. To make our life easier

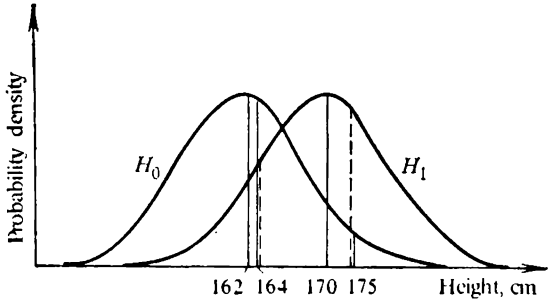


Fig. 2.

we will simplify the formulation of the alternative hypothesis: the mean height of the 1943-1948 age group is 170 centimetres, and hence over the 35 years it has increased by 8 centimetres. Figure 2 gives the probability densities for each hypothesis, the left curve describing the null hypothesis.

Let we have a sample of the later generation males. *A* first appears to be 164 centimetres tall. Such a height may be found in any group. Consider the dilemma: if the fellow is from the "left" group, his height is 164 centimetres with the probability density represented by the ordinate

value for 164 centimetres on the left curve (a solid line in the figure). If then he belongs to the group with the supposed new probability distribution given by the right plot, then he will be 164 centimetres tall with the probability density equal to the ordinate value for 164, but now on the right curve (a dash line). We have to make a decision—to assign the observed value to one of the two distributions. The suggested principle for decision making consists in giving preference to the distribution in which the observation is more probable. In our case, H_0 is more probable, and should this be the only observation, we should accept hypothesis H_0 (that height distribution did not change).

You will have noticed, of course, that the concept of probability has here been replaced by the probability density: it is easier to compare probability using their ratio and comparing it with unity. The ratio of probability densities is aptly called the *likelihood ratio*. It is compared with unity and according as it is more or less than unity, the "left" or "right" hypothesis is chosen, i.e. a more likely decision is made.

But it is not as simple as that. If a second fellow of the observed group is 175 centimetres tall, then reasoning along the same line we should assume the hypothesis that he belongs to the right distribution—here the right curve is higher, and hence more likely.

A comparison of the outcomes of the two observations is a problem: which is to be favoured. A good idea is to use two, or more, observations jointly. The "point" likelihood ratios are then multiplied together to obtain an expression,

which is a likelihood ratio too, but now for the whole set of observations. Next a number, called a threshold, is set, and the rule of accepting or rejecting hypothesis H_0 consists in comparing the likelihood ratio obtained with the threshold: if it is larger than the threshold, hypothesis H_0 (left curve) is accepted, otherwise hypothesis H_1 (right curve) is accepted. In the acceleration problem, in particular, the evidence is convincing: the 1943-1948 age group is on average 8 centimetres taller than the 1908-1913 age group.

As a matter of fact, the height problem does not necessitate the use of the criterion of likelihood ratio, since with the significant body of statistical data available we could do with simpler techniques. But there are problems in which this criterion is of much help and yields good results, for example, in the radar problem.

Noise, as is well known to experimentalists, obeys the normal distribution law with zero mean, and the signal plus noise also obey this distribution but with another mean—exactly the amplitude of the signal pulse. It is here that the likelihood ratio criterion is used to make the important decision as to whether there is signal and noise, or just noise without signal. And in general they make extensive use of this criterion in communication and control to test null hypotheses.

Maybe

To return to statistical tests of hypotheses, we can safely say that the figure of merit of test methods is the significance level: the lower the significance

level the better the test. But our reasoning so far has completely ignored the second-type error, which in rejection problems means the acceptance of a bad lot, and in the radar problem means

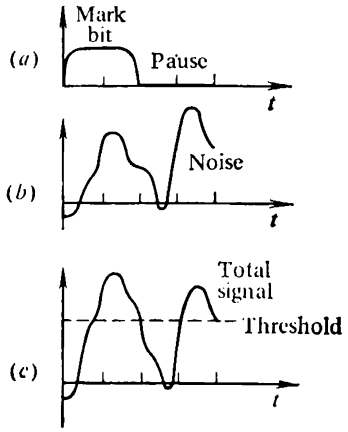


Fig. 3

the omission of a signal. Let its probability be β, then the measure of confidence for the statement "H_1 is true" is the probability $1 - \beta$.

The fact is that in the test method selected the probabilities of the first- and second-type errors appear to be dependent on each other, so that they cannot be specified in an arbitrary way. We will take an example from telegraphy, where the characters are mark pulses and pauses. Different situations are illustrated in Fig. 3. Noise distorts the signal, and the real signal arriving at the receiver is so obliterated that it is not at

all clear whether it is a mark pulse or a pause. The simplest way of processing such a signal in the receiver is to establish a threshold: if the signal as received is above the threshold then the decision is that a pulse has been received; if it is below the threshold, the decision is that there is a pause. In noisy channels the probabilities of errors of both kinds will depend on the value of the threshold: the lower the threshold the higher the probability that a mark pulse will be received correctly, i.e. the lower the significance level; the probability of second-type error will increase, though. In the test method described the selection of threshold is thus the selection of probabilities of errors of both kinds.

In radar and quality control the situations are similar. But how are the probabilities of the errors related? Suppose we want to provide a very small probability of false alarm in air defence, say, no more than one false alarm in ten million pulses, i.e. to select the significance level $\alpha = 10^{-7}$. You cannot, however, get the things like that for nothing—there is always a price to pay. If we were to bring the reasoning to the extreme, i.e. to make the significance level zero, we would have to treat all the signals as noise. Now the operator will never make false alarm, just as any alarm at all for that matter, even if a fleet of hostile aircraft will be overhead. This test procedure, normally called the plan, can hardly be considered satisfactory, to say the least.

This argument illustrates how necessary it is carefully to analyze the implications of first- and second-type errors and to arrive at some compromise.

It would appear that mathematical statistics should have worked out some ways of establishing or calculating the permissible probabilities of errors in hypothesis testing. But, alas, the permissible value of the significance level is a measure of risk taken by the decision-maker, viewing the situation from his point of view.

Later in the book we will look at the risk in more detail, but at the point it is worth noting that the permissible risk level is an extremely subjective thing. You will have played cards or some gamble and will know that the gambler's behaviour is dramatically dependent on the amount of possible win or loss and on the character of the game, its heat. And in problems concerned with engineering and nature the estimates of the permissible risk appear to be also dependent on a variety of other factors, such as prestige or qualification.

But the situation is not as hopeless as it might appear. So far we have only dealt with the simplest way of processing of incoming signals. When, speaking over telephone, you are poorly heard, you repeat the phrase several times. Also, in radar you can send out not one, but several pulses repeatedly, or a pulse packet, as it is normally called. Using a packet, we can now test the hypotheses by different information processing techniques.

We will mark 0 when we accept H_0 (noise) and 1 when we accept H_1 (signal and noise). Let the packet have 100 pulses, and for each of them we make a decision whether it is 0 or 1, i.e. the processed packet is 100 digits, each of which is 0 or 1. We can, for example, accept the

rule: if among 100 characters there are more than five 1's (and hence less than 95 0's), we will make the decision that a signal is present, otherwise we decide that there is noise alone. If instead of five we take any other number, we will have another test rule. In short, it is clear that the probabilities of errors of both kinds will vary with the number of pulses in a packet, the permissible number of 1's to make a decision that there is no signal, and, of course, with probability of correct detection of each pulse. We have, thus, many possibilities, but we should be extremely careful in selecting the procedure to be used.

When a friend invites you to take part in an outing, you may say YES, or NO. But you can also say MAYBE.

There is an infinite variety of motives for your selecting one or another of these answers, as is of implications. But notice the third of the possible answers, which differs markedly from the first two by its uncertainty. By saying MAYBE you put off the final decision. And you do this with good reason: you may be not acquainted with the other members of the party or have no idea of the route, you are uncertain about the expenses or not sure that you will get time off. And so you say MAYBE because you have not all the information required to make the final decision.

With statistical tests of hypotheses the situation is about the same: when the evidence available is not sufficient for decisive YES or NO, i.e. to accept or reject a hypothesis, we then can put off the decision by saying MAYBE and seek the required information.

The form of the radar problem just discussed in which pulse packets are used is not without its weaknesses. If the decision is made that there is a signal when a packet has more than five 1's and the first ten pulses contain seven 1's, then the decision will be made that a signal is present independently of the remaining 90 pulses. The processing of the 90 pulses will thus be a waste of time and power. On the other hand, if the first 70 pulses contain not a single 1, then we could with good reason believe that there is no signal and not to test the remaining 30 pulses.

This suggests that we might as well make the decision, i.e. accept or reject the null hypothesis, without waiting for the whole of the packet to be processed. So if 0's in the packet are few, it would pay to accept the hypothesis that there is a signal, and if 0's are many, it would be as reasonable to accept the null hypothesis. But when the number of 0's in the packet is neither very small nor very large, the information is not sufficient to reach a decision, and so instead of YES or NO we say MAYBE and make further observations.

The simplest plan of such a type consists of two successive packets.

Consider an example. In testing the no-signal hypothesis we can take a packet of 40 pulses, say. If we find not a single 1 or just one 1 among them, the rest being 0's, we accept the hypothesis that there is no signal; if six or more 1's, the hypothesis that there is a signal; and if from two to five 1's, another packet, say of 30 pulses, is beamed. Now we will have to set another permissible number of 1's, but for the total number of pulses,

i.e. for 70. We can here take it to be seven. And so if in the combined packet we will find less than seven 1's among the 70 pulses, we accept the hypothesis that there is no signal, and if seven or more, the alternative hypothesis. The quality control problem has the same structure.

Calculations indicate that such *double sampling plans* are more efficient than single sampling plans. A sample on average contains less pulses (or products) with the same or even better results.

Now you may well ask, why confine ourselves to two samples, and not make three, four, etc. samples?

This is exactly the case in reality. But it is one thing to have a vague insight, and quite another to work out a far-reaching theory.

Even with sampling control problems it is clear that a multitude of forms of plans are possible. And for the whole spectrum of problems of hypothesis testing a wide variety of rules can be developed, thus offering the decision-maker a wide scope.

Compromise

But which plan or rule is the best? As we know already, we should, above all, establish the criterion of quality for the test.

Let us begin by specifying the significance level (recall that it is the probability that the null hypothesis is rejected, although it is true), such that would satisfy us in this situation. Since the plans are many, the respective second-type errors will be many as well. Clearly, we wish to mini-

mize the probabilities of errors of both kinds. Therefore, for a given significance level we can use as the quality criterion the probability of the second-type error.

The problem at hand is called the *optimization* problem: given the significance level, select the rule for reaching a decision, such that the probability of the second-type error would be the lowest.

This principle of selecting the optimal rule lies at the root of one of the methods of testing hypotheses put forward by the outstanding American statisticians Neyman and Pearson in mid-1930s. The rule is based on the likelihood ratio.

The Neyman-Pearson criterion is widely used in statistics and, unlike the "essay" problem discussed earlier in the book, is couched in formal terms.

It is worth mentioning here that there are many other sound criteria. At the same time, in some cases the Neyman-Pearson criterion is rather vulnerable. We would like to discuss them now.

It would be instructive to look at the measure of confidence from another angle. We have already established that with the above test rule the price paid for the lower significance level is the higher probability of the second-type error, e.g. in sampling quality control we have losses due to rejection of good products.

And here, as it was said above, we have conflicting interests of the supplier and user. To arrive at a compromise they have to take into account all the possible consequences of the

first- and second-type errors. The parties to a deal consciously or subconsciously seek to minimize their losses.

In everyday life we have to adapt, to take into account the attitudes of top brass, subordinates, colleagues, the members of the family and neighbours, in other words, to live in society. Adaptation at best is consensus, but it may be that the interests of the sides are not only different, but even conflicting.

When on a Friday evening a couple are dressing for a visit and She has already dressed up, a difference emerges. He would like to go in his everyday suit and worn-in shoes: in this outfit it is more convenient to watch a soccer match on TV and have a cup of tea. But She insists on the black suit, starched shirt and patent-leather shoes—the hostess should be taught a lesson how to look after a husband properly. But the new shoes are not worn in and pinch a little, and the very idea of a stiff collar with a tie gives Him a pain in the neck. After much argument He puts on the new suit, a flanel shirt without a tie and old comfortable shoes. Another of the family crises has been worked out peacefully.

When the reconciled couple arrive at their friends' the shrewd hostess immediately take notice of the gorgeous looks of Hers and makes conjecture as to Her interest in one of the male guests.

The interests of the ladies here, although not coincident, are apparently not conflicting either. One is just curious and just wants to have a trump, the other, if the hypothesis is true, does not want a show-down..

Thus, the null hypothesis of the lady of the house is "the interest is there" The first-type error is to reject this hypothesis when there is an affair—it can make no harm to the hostess, if only a slight disappointment. But the second-type error—to accept the hypothesis when in actuality there is no affair—can start a dangerous gossip.

At the same time, if the hypothesis is true, She can foresee the consequences and take measures if the lady of the house smells the rat.

The compromise here is for Her to select so her behaviour that She could, for one thing, make use of the chance (or prepared) coincidence and, on the other, not to show Her cards.

The implications are moral, rather than material, and no participant at the party can establish a permissible quantitative measure for errors, or the cost of errors.

But the case of the supplier and user is clearly subject to quantitative analysis.

In real life, however, it is by no means easy to introduce a quantitative measure. The fact is that the permissible (in a given situation) probability of error must be specified in some reasonable way. But in which way—there is no answer so far. The situation appears to be rather complex.

Returning to the radar problem, recall that the detection system is very sophisticated and expensive, but its failure may be immeasurably more expensive. So a failure of a ship-borne radar may lead to a collision with another ship, iceberg or a cliff.

Now what probabilities of first- and second-type

errors would satisfy you? Suppose you chose to have 0.001 for each error. This would imply that if signals would be arriving, say, each 0.01 second (i.e. hundred signals a second), then the first-type error of 0.001 means that there will be on average one omission in each thousand signals, i.e. one in ten seconds and six in a minute. But these are average figures. But in actual practice there may simultaneously be several omissions as well. So such a detection system is good for nothing. If we set 0.000 001 for the same one hundred signals a second, then erroneous solutions will on average be reached once in ten thousand seconds, i.e. in about three hours. But we would like to have an absolutely error-free system, although it is well known that there are no such systems. What is to be done then?

And so the detection hardware is made ever more sophisticated and expensive, with large crews of skilled technicians providing adequate reliability of these systems.

Consider an example from everyday life. While crossing a street we run some risks even if we observe the rules conscientiously, because the rules may be violated by a motorist or a running boy who may accidentally push you under a car.

A daily average of injuries and deaths in traffic accidents in a Soviet town was two persons. If we take the frequency for the estimated probability—here unfortunately we have enough evidence—given that the town's population is 1 million, the probability of being involved in a traffic accident on any day for any inhabitant of the town will be 0.000 002, a figure not to be completely overlooked when it concerns your life.

Any mother will indignantly reject the very formulation of the question about the permissible (for her) probability of the first-type error, i.e. of an accident in which her beloved child may be killed. For mother the only permissible probability here is null. No use trying to explain to her that the only way of achieving this probability is to stay at home for ever.

It is thus unclear what principles we are to be guided with in selecting the permissible probability of the first-type error for accidents. And what principles should be used by local authorities in setting speed limits. Pedestrians may think that if, say, the limit is increased from 50 to 60 kilometres per hour, this will make accidents more likely. In practice, this is not necessarily so, since higher speed limits are as a rule accompanied by improvements in traffic control, population education, higher penalties, and so on.

But in reality, nobody sets this probability, and both motorists and pedestrians complain about the traffic police being unable to ensure adequate safety in the streets.

The main purpose of the traffic rules is to reduce the error probability. Here it is the first-type error with the generally accepted null hypothesis: in crossing a street no accident will occur with me. At the same time the false alarm, or second-type error, is here quite admissible, since it only implies that you may wait a bit being overcautious.

In our everyday life we cross streets, although we know that an accident is possible, because we are guided by the *practical confidence principle*: if the probability of an event is small, it should

be thought that in a single trial the event will not occur.

As early as 1845 P. Chebyshev wrote in his master thesis entitled *An Elementary Analysis of Probability Theory*: "Approximately, we consider it undoubtable that events will or will not occur if their probabilities are but slightly different from 1 or 0".

It is this principle that makes it possible for us to live without being constantly gripped by fear of accidents, which occur with such a small probability.

But what is to be understood by *small probability*? The same question: one hundredth or one millionth, or less?

For a normally functioning valve in your TV set, the probability that some electron flies from the cathode to the anode in a second is about 1/1,000,000,000. It would seem then that there should be no current in the valve. But the electrons are legion, and during a second about 10^{16} electrons come to the anode, and so the probability here does not appear to be so small.

The situation with accidents in a town is about the same. Although the probability for you personally to get involved in an accident is small, the population being large, the probability of at least one or even several accidents will be not at all small, and sometimes even close to unity, and so we often witness ambulance cars tearing along the streets.

Here too the notion of smallness in evaluating the probability is subjective. For teenagers, who generally overestimate their capabilities, the

threshold is overestimated as well. Besides, the estimation of probabilities may be different from situation to situation.

Consider an example of such a situation.

According to evidence available, in the USA the incidence of cancer of the stomach is lower than in Europe. This fact was attributed to the way of having drinks: Americans take hard liquors diluted (whisky and soda, gin and tonic, etc.), whereas Europeans have them straight.

Just imagine two groups of researchers, one preferring straight drinks and out to prove their wholesomeness, or at least that they are not as offensive as diluted, the other preferring diluted drinks and going overboard to prove exactly the opposite. What is more, the "diluted" group may champion the interests of the manufacturers of soda water and tonic.

The problem here is to test the null hypothesis: among the stomach carcinoma cases the share of "straight" drinkers is equal to that of "diluted" drinkers. The alternative hypothesis: among the cases the share of "straight" drinkers is larger than that of "diluted" drinkers.

It would seem that such a formulation of the problem should lead to a clear answer: the null hypothesis is either true or false. But this is not always so. The "straight" group seeks to prove that the null hypothesis is true, and therefore they would profit from such a test rule that would reject the null hypothesis only rarely, i.e. they would like to select the significance level as low as possible. They do not care for the second-type error, which here implies that a hypothesis is accepted that "both drinking habits are equally

offensive", when in reality straight drinks are more harmful.

At the same time, the "diluted" group are interested in the opposite interpretation of the same observations, and so they strive to have a smaller probability of the second-type error, i.e. to have the alternative hypothesis rejected as rarely as possible. Also, this group would like to have a large probability of rejecting the null hypothesis, if it is false.

Having so conflicting approaches and understanding of the importance of first- and second-type errors the groups may work out absolutely different rules of decision making, and hence obtain different answers.

The alternative hypotheses won out on statistical evidence and so the "diluted" groups, and hence the American drinking habit triumphed. But this did not put off the "straight", and they were successful in proving statistically that with the large intestine the situation is opposite. It might appear that this should have reconciled the sides, but today the surgery and therapy yield better results with the large intestine, and so the "diluted" gained an advantage.

But back to the Neyman-Pearson criterion. Some of the examples just considered would seem to suggest that this approach is not watertight at all. Neyman and Pearson criticized the arbitrary selection of one of the two alternative hypotheses, as was the case before them. Their theory, however, just transfers the arbitrariness to the selection of the permissible significance level. To be sure, after the level has been chosen the hypotheses are further tested with all the

adequate rigour. So the arbitrariness is still there, but it is only hidden deeper, so that now it is more difficult to perceive the implications of it.

This is not to suggest that we should discard the Neyman-Pearson criterion. There are problems where we can quite reasonably establish the significance level. What is more, in some cases we can even derive the analytical dependence between the probabilities of the first- and second-type errors and clearly see the price to be paid for reducing the level.

Although the Neyman-Pearson theory was a major breakthrough, it did not give all the answers. Even now, after decades of intensive studies, the theory of statistical testing of hypotheses is still far from its conclusion.

Dynamics Instead of Statics

Let us once more return to the radar problem. It is easy to criticize and it is not surprising that the test rule was readily dethroned above, in which a pulse packet of a fixed volume is used. A two-packet plan is better, but as you may have noticed the procedure can be pursued further, i.e. we can use three- and four-packet plans, but they have one essential drawback, the staticity.

Such an approach contradicts our everyday experience. So before taking a decision when buying a new suit, changing position, etc., you weigh all the pros and contras, and as a rule a priori do not fix the time or the number of observations sufficient to make the decision.

Why then in the case of quality control, radar, or other problems of those discussed above should we follow the other procedure, fixing beforehand the sample volume? If in testing we preliminarily fix the sample volume, then the sequence of operations will be absolutely independent of the gradually accumulating data.

It is thus seen that the single sample plan is not reasonable. In double sample control the procedure is but slightly dependent on the results of the first sample (a repeated sample is not always necessary).

Therefore, the idea of *sequential analysis* suggested itself. Its main difference from fixed sampling lies precisely in the fact that the very number of observations is not fixed beforehand—it is only dependent at each stage of observation on the previous results and is, thus, a random variable.

It would seem that the simple idea not to fix the time or number of observations but to make decisions after the required information is available is apparent, and it is unclear why it has been overlooked for so long. Although, if we take a closer look at it, it will not be that apparent.

To begin with, it is only after the meticulous calculation that the advantages of the sequential plan show up in comparison with the single sample or more effective double sample plans. But the main thing here is that the dynamic analysis is a revolutionary approach.

The double sample plan was developed late in the 1920s. But it was only in the 1940s that the multistage samples began to be studied along with some of sequential procedures. But the

main breakthrough here came from the American mathematician Abraham Wald (1947).

Almost all major discoveries are simple, but they require a nontrivial, revolutionary way of looking at known facts, and so take a stroke of genius and even tenacity, since a fresh idea is always hampered by old concepts and prejudices.

To illustrate the Wald sequential test, we will consider a ballerina maintaining her weight. For simplicity we will consider the situation when there are two hypotheses: H_0—the ballerina's weight is 50 kilogrammes, and H_1—the weight is 48 kilogrammes.

Suppose that the dancing company goes on a tour, where the routine will be disturbed, but the dancer must still maintain her normal 48 kilogrammes. And so the weight fluctuations must be closely watched so that the diet and rehearsal habits might be changed in case of need. But the human weight is by no means stable, it varies within hundreds of grammes even during the day, and also it is measured with uncertainty. Therefore, the dancer here should weigh herself several times a day and watch the average value. But how often? Two or three times may be not enough, but a hundred is impossible. It is here that the sequential test comes in.

Remember the idea of likelihood ratio? A slightly modified form of it will now be used here. To begin with, we set the probabilities of errors of the first and second kind (which here can be assumed to be equal), it would be reasonable to set two thresholds and follow the rule: if the likelihood ratio is higher than the larger of the

thresholds, we accept the null hypothesis—the weight is 50 kilogrammes; if less, the alternative hypothesis; and lastly if tbe ratio value lies between the thresholds, there is no reason for accepting any of the hypotheses and observations must be carried on.

Instead of speaking about testing the null hypothesis or the alternative hypothesis and accepting one of them we can speak about rejecting the null hypothesis. Using this language we can say YES, if we accept the null hypothesis, and NO, if we reject it. But if we cannot say YES or NO (information available is not enough), we can say MAYBE and go on with our observations.

The situation can be presented graphically. The sequential values of the measured weight of the dancer will be denoted by x_1, x_2, . . ., x_n, the likelihood ratio being a function of these variables.

We can then assume that the human weight is distributed following the normal law. But if the null hypothesis is true, then the expectation for the weight will be 50 kilogrammes, and 48 kilogrammes for the alternative hypothesis. Under these conditions, if we want to find whether the likelihood ratio lies within or beyond the thresholds, we will have to test the following simple inequalities:

$$a + 49n \leqslant \sum_{i=1}^{n} x_i \leqslant 49n + b,$$

where n is the number of observations. The number 49 here is a half-sum of 50 and 48, and a and b

are governed by the predetermined probabilities of first- and second-type errors.

If now we denote the half-sum by μ, then

$$a + n\mu \leqslant \sum_{i=1}^{n} x_i \leqslant n\mu + b.$$

Figure 4 is a plot of these inequalities: the right- and left-hand sides are straight lines in the coordinates (n, x), and $\sum x_i$ is a broken line. When the broken line intersects the upper straight line, it will get into the region where the null hypothesis is valid, and the decision is YES, when it crosses the lower straight line, it will get into the region where the null hypothesis is rejected, and the decision is NO. And until, with increasing number of observations, the broken line varies within the confines, the decision at each step is MAYBE, and the next observation is made.

Turning again to quality control in mass production, when an automatic machine is operating and the parameter to be controlled is the product size, which must be strictly within the tolerance, the situation absolutely coincides with the one just discussed. What is only required here is to substitute the word "product" for "dancer", size for weight, and the null hypothesis now will be the exceeding of tolerances. But the exceeding of tolerance is "no go" here.

When the probabilities of errors of the first and second kind are similar, then, as was shown by Wald, the sequential procedure enables the number of tests to be reduced by half as compared with the earlier rule, where the number of tests is predetermined. In quality control, halv-

ing the number of products to be tested is a substantial gain, especially so if the tests are expensive and tedious. Therefore, the sequential analysis is widely used in quality control.

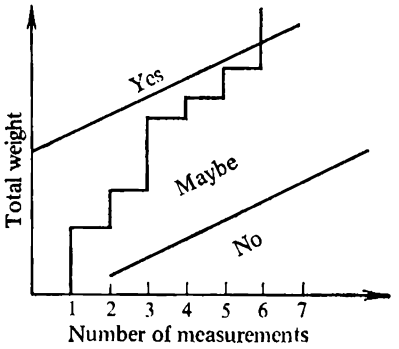


Fig. 4

There is also another set of problems where the sequential analysis is justified and advantageous.

In the radar problem higher reliability is achieved by repeatedly beaming signals and summing the received signals together. In the process, irregular noise signals, which irregularly assume either positive or negative values, partially cancel out, thus improving the signal-to-noise ratio.

Recall that in radar the two kinds of errors have costs that are far from similar. And if we assume that they differ drastically, then the sequential procedure in signal processing will provide a substantial gain in the number of

observations and also in the time and power required, as compared with the earlier procedure with a fixed number of observations. In urgent situations a reduction in observation time is a great advantage, and reduced power required amounts to increasing the range of detection. It is worth noting here that observations are fewer on average only, the spread being fairly large. Therefore, sequential analysis should only be applied after careful calculations, which lie beyond the scope of the book.

Risk

Two young friends want to go to the seaside for vacation. One of them owns a car, the other a motor-cycle. Besides, they can travel by train or air. Motoring has the advantage of freedom, but the car needs some repair but the time is pressing. As to the motorcycle, it is not the most comfortable vehicle to cover the 1000 kilometres to the seaside. On the other hand, with the car or motor-cycle the holidays start immediately, at the threshold. But there are troubles: first of all, an accident is more probable on a motor-cycle than in a car, train, or airplane. Also, to meet with an accident in a car is for the most part to have the car body damaged, whereas in a train you at best get away with fractures and bruises, and with the air travel... it is clear. And so the friends start a lengthy discussion of pros and cons of each option. Let us now approach the situation in scientific terms.

To have a good vacation is an understandable desire, and realization of this desire is a success, which we will term *win* or *gain*.

In the situation at hand the travel troubles and accidents are statistically stable events, and we can assume with good reasons that there are definite probabilities of meeting with an accident for any of the vehicles discussed. But the boys are interested not only in the probabilities of accidents, but also in the respective losses.

In an air crash the losses almost certainly are the death, i.e. extremely high. Another loss in air travel is the high cost of tickets. The gain—or negative loss—is two extra days spent on the seaside.

In a train accident the losses are death or injury, long stay at a hospital, travel expenses plus two days less spent on the beach.

Motoring takes much time, effort and funds in repairing the car; an accident may land you in a hospital or kill you, and leads to direct losses of time, money and effort. Travel expenses here are also high: petrol, lodgings, and so on.

Motorcycling is, in addition, extremely uncomfortable for long-distance travel.

Some of the losses can be expressed in terms of numbers (hours, money, etc.), the others cannot—you can hardly work out a price of a fracture of the hand, pelvis, or concussion of the brain, and less so of the loss of your life. There are some gains too: you show your mettle by fighting through the difficulties, the comfort of having a vehicle on the vacation, and other moral benefits. Now to arrive at the best decision we will have to work out the total loss.

The answer is by no means easy. Above all, we must be able to compare all the losses. For this purpose, the best idea would be to express them in the same terms, say, to estimate the money worth of the time wasted, the moral loss and possible injuries. Some hints might be given here, but it would lead us too much astray. A further example will illustrate the line of reasoning.

When deciding whether or not to insure your car against stealing you may draw on mathematical statistics. During the period of insurance the following two outcomes are possible: ω_0—the car has not been stolen, and ω_1—the car has been stolen. The two events can be considered random (at any rate from the point of view of the owner), and probabilities of any of the two outcomes can be estimated from the police statistics: $P(\omega_0) = p$ is the probability that your car will not be stolen during the year, and $P(\omega_1) = q = 1 - p$ is the probability that your car will be stolen during the same period.

You may make two decisions: d_0—you pay the insurance premium (insure your car), and d_1—you do not.

Your possible losses here will in some way or other involve the insurance premium (r_0) and the car price (r_1). Suppose, for simplicity, that the insurance indemnity equals the car price.

Now we would like to work out losses in various situations. The losses in the case of the decision d_j and the outcome ω_i are conventionally designated by $L(\omega_i, d_j)$. So $L(\omega_0, d_0)$ are the losses for the case where the car has not been stolen during the year and you paid the premium. Here

$L(\omega_0, d_0) = r_0$, i.e. the value of the premium. Similarly, $L(\omega_0, d_1) = 0$, since the car has not been stolen and you paid nothing; $L(\omega_1, d_0)$ is the situation where you insured your car, it was stolen and hence you will be returned the car's price, your loss here being the premium alone: $L(\omega_1, d_0) = r_0$.

Lastly, $L(\omega_1, d_1)$ is the situation where the uninsured car was stolen, so that your loss is the car's price: $L(\omega_1, d_1) = r_1$.

The risk can be estimated by averaging losses in all the situations possible. The most natural measure of the average loss is, of course, its mathematical expectation for the decision d, i.e.

$$\rho(d) = \mathsf{M}L(\omega, d),$$

where M is the sign of mathematical expectation.

The quantity M is called the *risk of deciding d*. We will now proceed to calculate it for both cases possible.

If you make the decision d_0, i.e. insure the car, then

$$\rho(d_0) \quad L(\omega_0, d_0)\, p + L(\omega_1, d_0)\, q = r_0 p + r_0 q = r_0 (p + q) = r_0.$$

Thus, if you insured your car, whatever the outcome, your risk is just the value of the premium.

If you chose d_1, then

$$\rho(d_1) \quad L(\omega_0, d_1)\, p + L(\omega_1, d_1)\, q = r_1 q.$$

This illustrates the meaning of the concepts "large risk" and "small risk": if at d_0 the risk is much less than at d_1 (i.e. in our case, if $r_0 \ll r_1 q$), then you should not take d_1—the chances of heavy losses are too high. Should the relation-

ship between the risks be the opposite, it would make sense to take risks—to do without the insurance and rely on chance.

Let us now make a rude calculation. Suppose, as is often the case, your premium is a preset percentage of the property to be insured. Let r_0 be 1.5 per cent of r_1. Then instead of $\rho(d_0) \ll \ll \rho(d_1)$, or in our case $r_0 \ll r_1 q$, we have r_0, using the relationship $r_0 = 1.5 \cdot 10^{-2} r_1$,

$$1.5 \cdot 10^{-2} r_1 \ll r_1 q$$

or

$$1.5 \cdot 10^{-2} \ll q.$$

So if the above relationship holds, then the probability of stealing is larger than 0.015, and you should better insure your car. If, however, $0.015 > q$, then you should not insure your car, since on the average you would suffer a loss.

The procedure just described does not include, of course, the moral aspect. Having your car stolen is bound to give you a stress, whose implications are now dependent on your individual capability to cope with stress situations: you may just grieve a while or get a myocardial infarction. For such additional factors to be taken into account we will need additional, more stringent requirements for the acceptable probability of stealing.

We have only considered the insurance against stealing. But insurance policy normally provides compensation at damages. This complicates the problem, although the reasoning here follows the same lines: we will have to take into account a wide variety of possible outcomes of accidents

with different degree of damage considering not only the repair costs but also possible injuries, even the lethal outcome.

An experienced or shrewd motorist knows the frequency (or probability) of these accidents. These probabilities are known as a priori probabilities: they are known, or rather assumed, before the observation. A knowledge of a priori probabilities makes it possible to work out the expectation, or the average risk, i.e. to average the risk over the a priori probabilities.

The average risk for the motorist to meet with a disaster, i.e. to have his car stolen or crashed, will thus be

$$\overline{\rho} \quad g\rho_1 + (1 + g)\,\rho_2,$$

where ρ_1 and ρ_2 are the risk of stealing and accident, respectively, g and $1 - g$ are a priori probabilities of stealing and accident given that one or the other, unfortunately, occurred (hence the sum is unity).

The risk averaged over a priori probabilities, and hence the decision made that provides the minimum risk, is termed Bayes's risk.

The Bayes approach is convenient, since it assigns to each decision a number, so simplifying the search for the optimum decision. The approach is therefore widely used in decision theory. But it is not without its weaknesses: it is based on a knowledge of a priori probabilities. There are situations where a priori probabilities can be thought of as known, as in quality control of mass-produced articles or in some of the problems of diagnosis in medicine. But unfortunately there are many situations where a priori probabilities

are not only unknown, but even senseless. So in the radar problem it would make no sense to speak about an a priori probability for a hostile aircraft to be detected in the region. Accordingly, in cases the a priori probabilities are unknown or their specification is extremely difficult the Bayes approach must be rejected and other methods, such as likelihood ratio of Neyman-Pearson, used.

The Picky Bride Strategy

Any girl is looking for the Prince among her admirers, and thinks twice before going to the City Hall with the chosen one for the wedding license. The task is not all that easy, otherwise we would not witness the divorce binge these days on the initiative of young women.

Nay, I do not want to discuss the various aspects of the complicated and exciting problem, but I will only try to consider one of its mathematical models.

Let our beauty have n grooms, encountered in an arbitrary order. To be sure, it is next to impossible to quantify the virtues of a groom, especially for a girl, for whom often the main characteristics are such human values as love, elegance, virility, etc., that are not amenable to formalization. However picky and choosy, the girl may eventually make her decision.

We will consider the following groom selection procedure: the bride meets her grooms in succession and after this (the period of acquaintance is immaterial) she can either turn him down or

call him her chosen one. Three important conditions are assumed to be met: first, the grooms turn up in succession (no two appear simultaneously); second, she does not come back to the groom she has turned down; third, the number of grooms is preset.

These conditions may well appear unreal to you: in everyday life a girl can sometimes return to the rejected suitor or date simultaneously with several boys. Before we set out to work the problem in mathematical terms, consider another possible problem.

In a machine-building factory the head of the first shop was offered the privilege of selecting five workers among 30 young fitters fresh from the vocational school. But as he rejects a boy he never gets him again because the personnel department immediately sends him to other shops. The manager is critically interested in having the best hands. What strategy is to be followed? The young fitters come to the shop one by one, and after an interview, acquaintance with the records and a test job the manager has either to accept or to turn down the candidate. Our mathematical model also assumes that the manager not only knows the total number of the applicants (in our case 30) but he can choose the better one among any two.

Put another way, the main feature of the problem at hand is the ability to order objects according to their quality, or to use the language of decision theory, according to preference among the objects already considered.

The problem of choosing several workers (m) among the available group (n) is a generalization

of the previous one, where only one, the best, is selected, i.e. where $m = 1$.

Let us now formalize the problem statement for $m = 1$, and n objects ordered by quality. The quality characteristic can be represented by a number or point (a) on a real axis: the higher the quality, the larger the number or the more to the right lies the point.

Objects are considered in a perfectly random order, so that the coordinate a_1 of the object that appears first may with equal probability be any of the n points available. Similarly, a second object with coordinate a_2 may with equal probability be any of the remaining $n - 1$ points. Sequential observations will thus yield a set of coordinates $a_{i_1}, a_{i_2}, \ldots a_{i_n}$, each of their possible $n!$ permutations being equiprobable.

Points (or objects) occur sequentially, but our task here is to stop once a point with the largest possible coordinate turns up, thus selecting the object with this coordinate and making no more observations.

But in reality we do not know with complete certainty that the object at hand has the largest coordinate, since we can only compare it with those observed earlier, not all of them, and so the only way to be dead sure that we are right is to meet the best object at the last step. The situation can, of course, be tackled by the theory of probability. It will therefore be a good idea to state the problem in this way: find a path leading to the right decision with the highest probability.

What strategies are available here? We can stop at the very first step, i.e. select a point with

coordinate a_1. For the picky bride this implies selecting the first groom who turns up. Using this strategy, she can at once put on the wedding ring, but the probability of her having the best candidate will only be $1/n$. If the claimants are legion, i.e. n is large, in that case the probability of getting the best is quite low.

It would seem that with any strategy the probability of choosing the best groom or, in the more formal statement, of a_i being the largest coordinate, will fall off indefinitely with increasing n. This, however, is not the case.

Let n be even. We will follow the strategy: skip the first $n/2$ points and then choose the first point with coordinate larger than that of any earlier point. Calculations show that the probability of hitting upon the largest coordinate in this strategy will be more than 0.25 whatever the value of n.

We have thus a strategy leading to success with an appreciable probability. Since n is fixed, there exists an optimal strategy providing success with the highest probability possible, which consists in the following: a certain number s of objects are skipped, the first object better than any of the previous is chosen. We can find s from the double inequality

$$\frac{1}{s+1}+\frac{1}{s+2}+\quad+\frac{1}{n-1}\leqslant 1<\frac{1}{s}+\frac{1}{s+1}+\quad+\frac{1}{n-1}.$$

The probability of having the best object will thus be

$$p_n=\frac{s}{n}\left(\frac{1}{s}+\frac{1}{s+1}+\ldots+\frac{1}{n-1}\right).$$

For example, our beauty has to choose among 10 grooms. From the last expression we can readily have for $n = 10$: $s = 3$ and hence the optimum strategy for the picky bride will be to ignore the first three grooms and then select the first who will be better than any of the previous claimants. The probability of the best selection here will be $p_{10} \approx 0.4$.

If n is very large ($n \to \infty$), this strategy gives

$$p_n \approx \frac{1}{e} \approx 0.37,$$

where e is the base of natural logarithms.

Quality Management

Many think of James Watt (1736-1819) as the founder of automatic control. He developed the centrifugal governor for the steam engine. The prototype steam engine took Watt much efforts. An American, William Pies, writes: "The first machine in the modern sense of the word was an appliance for boring cylinders invented by John Wilkinson in 1774. Wilkinson is not known so widely as Watt, although it is his invention that enabled Watt to construct a functioning steam engine. For a decade Watt was making futile attempts to manufacture a cylinder to an accuracy required. After one of his attempts he said in despair that in his 18-in. cylinder 'in the worst place the deviation from cylindricity is about 3/8 in.' However, as early as 1776 Watt's assistant Mathew Bolton wrote; 'Mr Wilkinson bored

for us several cylinders almost without error: for a 50-in. cylinder we then installed in Tipton deviations were never larger than the thickness of an old shilling.'"

Now I cannot estimate exactly the percentage error—I do not know the thickness of a shilling that had already been old late in the 18th century, but indications are that the accuracy had improved several times. In consequence, Wilkinson's cylinder boring machine made Watt's steam engine commercially viable, and so the machine was direct ancestor of modern precision metal-cutting machine tools.

In my opinion a major advance in the development of industry was the advent of interchangeable components. It would hardly seem reasonable nowadays to produce each detail as a unique piece and then adjust them to one another. But when in 1789 Ally Witney organized production of muskets on order of the American government based on his idea of assembling muskets from interchangeable details most of the experts of the time were suspicious and dismissed the idea as being of no practical value.

Our today's industry turns out millions of high-accuracy interchangeable components, so not only doing away with the need for expensive manual labour but also achieving the standards of accuracy far beyond human powers. And still we are not always satisfied with the quality of production. Here we will deal not with acceptance control but with the issues of quality management during the manufacture stage. This management is based on the routine in-process check. To be more specific, we will discuss an automatic

machine producing, say, bolts, where the parameter to be controlled is the bolt length.

The control consists in measuring the length of sampled bolts and comparing the measurement results with a preset value. Sure, the measurement results do not always agree with the reference, exhibiting slight deviations from the average due to misalignments in the machine, power supply fluctuations, inhomogeneities of the material, and so on. These deviations are a typical example of a random variable with a continuous probability distribution $p_0(x)$.

But in case of some malfunction in the machine there occurs either a noticeable deviation of the average value of the controlled parameter from the reference, or a noticeable increase in spread, or both. In other words, the probability distribution of bolt length changes. Denote the probability density of the new distribution by $p_1(x)$.

We thus have two hypotheses: H_0—the machine is functioning normally, and H_1—a malfunction, the null hypothesis having the probability density $p_0(x)$ and the alternative hypothesis $p_1(x)$.

A malfunction results in rejection, and so the task of quality management here is to determine the moment, as quickly as possible, when the malfunction has started, then to stop the machine and to remove the problem.

As in the case of a picky bride, we should optimize stopping the machine. We can also note some difference in the problem statement, however.

Reject—bad bolts—is loss, an entity readily expressible in monetary terms. Adjustments of the machine mean downtime, i.e. again loss

expressed in monetary terms, which enables us to formulate an optimization problem: to stop the machine so that the loss due to downtime would be minimal. In that case the errors of the first and second kind are respectively the production of some reject due to miscalculated stop, and false alarm, i.e. shutdown when the machine works normally. Losses due to each error and average losses can be readily expressed in monetary terms.

At the same time, as we have already discussed in the previous section, search for the best groom, or a new worker is not that easily amenable to quantification of the losses due to miscalculated stop (i.e. erroneous selection of the stop point). Therefore, in the previous section we did not pose the problem of optimization of the last observation point providing the minimum risk. Instead, we confined ourselves to the strategy ensuring selection of the best object with the highest probability.

The last observation point both in the quality management and bride problems can be easily found from the Bayes relation.

The Mathematical Model

A children's drawing showing a house with a chimney and smoke, the round sun with rays, fur-trees and ox-eye daisies is a model of the surrounding world, illustrating the main elements of the perception of children.

An artist, realist or impressionist, will paint the same scene differently. Depending on his

philosophy the painter looks for certain aspects in the scene around him that he can render.

But even an orthodox naturalist is not able completely, with absolute accuracy, to reproduce nature, even if he were in a position to reproduce everything he sees, because he would not be able to render motion, smells, sounds, all the variety of life.

A model is normally simpler than the object it models. By the way, a girl sitting for a stature is also called a model. In that case, a statue of the girl will hardly be more complex, speaking in human terms, than the girl herself, but this terminology seems to reflect the relationship of the sculptor's plot, generally rather deep, with the appearance of the model who is here only a material helping to incarnate his plot.

A drawing of a machine or a schematic diagram of a radio receiver are models too, and it is not surprising that in radio design they pass successively from block diagram through schematic diagram to wiring diagram. All these model the future set not only in varying detail, but they also reflect its different aspects. So the block diagram describes operations performed by the component blocks of this piece of radio equipment, the wiring diagram indicates the connections of the components—resistors, capacitors, transistors, etc.—to perform their specific operations. But after the radio set has been manufactured, it is debugged: they reduce capacitance here or increase resistance there. The wiring diagram is a model, not the set in itself.

In recent years the notion of model has been used widely in a variety of scientific disciplines,

in engineering, arts and even in fiction. It seems to be impossible to describe, and even more so to define this concept, which would formally correspond to all its applications and at the same time would be understandable to the members of so different professions. But for us in what follows it is not as necessary.

We will here make use of a narrower concept of mathematical model—description of an object studied using a formal language, i.e. numbers, equations (finite, differential, integral, integro-differential, operator equations), inequalities, or logical relations.

Population growth in a town is proportional to the number of inhabitants. A mathematical model here is a linear equation and it is only valid in a fairly rough approximation. If we take into account the old people, children, unmarried women, the model will become much more complex. And if we include such factors as education, employment of women, income levels, etc., the model will become so complex that its construction and investigation will present enormous difficulties. But even all those factors included, the model may turn out to be still a far cry from reality, since it ignores a sea of random factors, such as population migration, marriage and divorce statistics, and so on and so forth.

Let us now consider the process of petrol manufacture from crude oil. In primary fractionation, petrol is obtained by vaporization: oil is heated to a certain temperature and lighter petrol fractions are vaporized to be removed at the top of the fractionating column. If the temperature at the bottom of the column and that of the crude

oil at the input are taken to be constant, then the simple mathematical model relating the amount of petrol derived to the amount of crude oil fed will be a linear equation: if we increase the input of crude oil 1.2 times, the petrol output will increase 1.2 times as well. Such is the simple and very crude model. If then we include the relationship between the crude oil input and the temperature and pressure in the column, the model becomes yet more complex. If we note that the column cannot accept just any more inputs and the temperature cannot be increased indefinitely (technological limitations) the mathematical description becomes more involved, but this is not the whole story: a real fractionation unit has about 200 measured, automatically controlled parameters, some of them with complicated feedbacks. Even if we write all such constraints and dependences, the resultant model will be of no value for control even with more advanced computer technology than we have now. In addition, there are some poorly controllable factors: random variations of qualitative characteristics of the crude oil at the input, random fluctuations of temperature, pressure, electric power supply, and so forth. Perhaps the modeller should give up the problem as a bad job?

The picture will be more impressive if we consider an attempt to produce a mathematical model of a living thing. Is it possible to model, say, the functioning of the brain of the dog or man, such that would reflect not only the activities of several billions of nerve cells, but also the interplay between them? Hardly so. On the

other hand, if we do not cover the variety of processes in the brain and construct a model involving only a small fraction of variables and relations, such a model will hardly satisfy us, since it will not give an adequate description of the situation.

Thus the term "mathematical model" is here viewed in a wide context. In his book *The Probabilistic Model of Language* V. V. Nalimov writes:

"Now we frequently read something different into the term 'mathematical modelling', taking it to mean some simplifications and approximate mathematical descriptions of a complex system. The word 'model' is here set off against a law of nature that is assumed to describe some phenomenon in some 'absolute' way. One and the same complex system may be described by different models, each of which reflecting only some aspect of the system at hand. This is, if you wish, a glance at a complex system in some definite, and definitely narrow aspect. In that case, understandably, there is no discrimination problem—a variety of models can exist concurrently. Viewed in this light, a model behaves in a sense just like the system it describes, and in another sense otherwise, because it is not identical with the system. Using linguistic terminology, we may say that the mathematical model is simply a metaphor."

What has metaphoric language to do with complex systems, such as an oil refinery or the brain? Why construct models then if it involves enormous difficulties of scientific, managerial and psychological character?

This is by no means a rhetorical question and we will look at it in more detail. Mathematical models are not constructed just for fun or to expand the list of scientific publications. At the same time modelling is a major (maybe the only) way of questing for knowledge, only not arbitrary modelling but such that provides an insight into some interesting and useful facets of a phenomenon under investigation, maybe ignoring some aspects of minor importance. If viewed from another angle these latter may turn out to be more important, and then another model will be necessary.

All the prejudices from superstitions of savages to fortune-telling from tea-leaves are based on misunderstanding of cause-effect relations and wrongly constructed models on which the predictions are founded.

At each step we construct some (nonmathematical) models: take along an umbrella or put on a light dress after having glanced at the heavens, step aside when we meet a lone dog, and so on.

When in trouble we also construct models: a girl sees her groom conversing with some blonde, she imagines faithlessness (another model) and in despair may either think of committing suicide (a model) or revenge (a model) depending on her character. In a split second she recognizes the blonde (the wife of her groom's boss) and her despair is gone.

So despair and hope are inherent in a model of a situation, which is necessary for a man to predict the outcome of events, and hence to select appropriate behaviour.

But back to mathematical modelling. Computations of planetary orbits—the modelling of their motion—is necessary to predit their behaviour: their rise and set, eclipses, and so on. Mathematical modelling gained acceptance with the advent of computers, which allowed one to tackle even complex systems—also called diffuse or poorly organized. In well-organized systems processes or phenomena of one nature can be singled out, which are dependent on a small number of variables, i.e. systems with finite, and small, numbers of degrees of freedom. At the same time, in complex systems it is not possible to distinguish between the effects of variables of various nature. For example, in the primary processing of crude oil we cannot separate the effects of material flows, temperatures, pressures in the various sections of a huge installation consisting of several 40-metre fractionating columns containing dozens of trays, banks of heat-exchangers and other sophisticated hardware.

This notwithstanding, we can construct mathematical models of such processes or objects to predict their behaviour and to control them. It appears that it is by no means necessary to include into the model all the variables—sometimes, if a problem is adequately stated, only a small number will be enough.

A mathematical distribution as such is a mathematical model too. Take for example one of the most remarkable—the Bernoulli or binomial distribution: if the outcome of independent experiments may be treated either as success or failure, the probability of success being constant and equal to p, then the probability of m

successes for trials will be
$P_n(m) = C_n^m p^m (1-p)^{n-m}$,

where $C_n^m \quad \frac{n(n-1)\ldots(n-m+1)}{n!}$ is the number of combinations of m elements taken n at a time.

The Bernoulli distribution is a good mathematical model of such events as gas density fluctuations, calls at a telephone exchange during a period of time, shot effect in a vacuum, current fluctuations in a valve, intensity fluctuations in oscillation composition, and unidimensional random walk, the simplest form of the Brownian motion. To be sure, the problem of coin tossing, with which traditional texts on probability normally begin, is also covered by the model. What is perhaps more important here is that this distribution, subject to some constraints, yields the Poisson distribution, and, subject to the others, yields the normal distribution, i.e. two models that are very common in applications. I think you are acquainted with these distributions, which are to be found in any probability course. Recall that the normal distribution is a mathematical model of experimental errors, height and weight of animals of one species and sex, etc. whereas the Poisson distribution is a model of radioactive decay, number of fires in a town, meteorite falls, railway disasters, and many other similar processes, often called "rare events"

In the literature devoted to applications of probability the role and versatility of these distributions seems to be overestimated. Books

addressed to practical engineers make one believe that no other distributions are necessary in nature and technology, since these well-studied distributions and their derivations, such as the log-normal distribution, seem to take care of all the situations. This, however, is not so, and in the ensuring sections I am going to provide examples illustrating the limitations of the above distributions as models for real problems.

Birth Statistics

Girls sometimes complain that there are not many boys at dances and discotheques. This cannot be generally attributed to the fact that there are less males than females. Maybe boys prefer football, wind-surfing or some other pas-times. Soviet statistics, for example, indicated that in 1982 up to 24 years of age there were 10.5 males per 10 females, i.e. 105 boys for every 100 females. This is not always the case. Wars take the largest toll of male population, and so in post-war years the sex ratio changes dramatically, thus giving rise to situations where there may be only 50 males for every 100 females.

Now we would like to discuss biological laws governing the sex ratio in the animal world and the mechanisms involved.

By definition, the sex ratio in a population is the number of males for every 100 females. Primary sex ratio obtains at conception, secondary sex ratio at birth, and tertiary sex ratio at sexual maturity. Various species of animals show

appreciable fluctuations of secondary sex ratio in various periods, which forms a subject of prolonged discussion in the literature.

Let us take a closer look at the secondary sex ratio in man, i.e. the number of new-born boys for 100 new-born girls. Normally this number is 105 to 106. The probability for a boy to be born is thus

$$105/(100 + 105) = 0.512,$$

i.e. slightly more than a half.

But fairly long periods of time see noticeable deviations from this pattern. These deviations and their causes are treated statistically in a large body of literature. So it was established statistically that during and after long wars in the countries taking part in the war the secondary sex ratio increases. For example, in Germany during World War I it reached 108.5, during World War II in Great Britain and France it increased 1.5-2 per cent. Figure 5 illustrates the growth of it for Germany.

Many hypotheses have been put forward to account for this. It was found that the secondary sex ratio increases for younger fathers, and during a war newlyweds are generally younger and some authors believe that this explains the situation. Others attribute this to the larger number of mothers having their first babies, for which the percentage of boys is higher. We will not here discuss all the hypotheses suggested, but it is only worth mentioning that neither gives a complete explanation of the results obtained. And if we want to find some clues, we will have to work out a model of the mechanism involved,

such that would be able to account for the way in which nature controls the reproduction as it suits it.

Nevertheless, there is no saying a priori that a woman will give birth to a boy, say, and therefore the probabilistic model here is the above-mentioned binomial distribution for the probability of having a boy other than 0.5. In a first

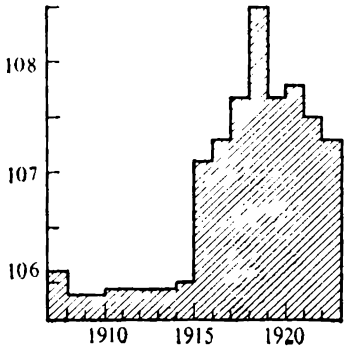


Fig. 5

approximation such model agrees with the observations, but a more rigorous test reveals its inadequacies.

If the binomial distribution were an adequate model of birth-rate in families, we could then predict the frequencies of any number and sex ratio of children. Thorough examination indicates that families of monosexual (only boys or only girls) or nearly monosexual (predomination of boys or girls) children are more frequent than

could be expected, assuming that the sex ratio is purely random and governed by binomial distribution.

Table 3 provides evidence, which appears to be fairly convincing. Figure 6 compares observed

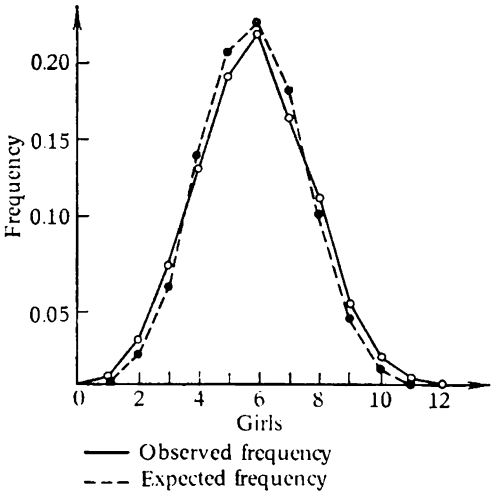


Fig. 6

and expected frequencies. The comparison shows that the binomial distribution here is not a good model. This can also be shown by using, for example, standard statistical criteria.

An attempt to produce an adequate model was made in 1965 by the Soviet worker V Geodakyan. He assumed that there exists some feedback (negative feedback in the language of wave

theory and cybernetics), which controls the secondary sex ratio should there be some deviations in the tertiary sex ratio, i.e. one at sexual

Table 3

Sex-ratio Frequencies in Families with 12 Children (5 million births)

Boys/girls	Observed total	Observed frequency per million	Expected frequency (binomial law) per million	Difference sign
12/0	7	0.0007	0.0003	+
11/1	60	0.0056	0.0039	+
10/2	298	0.0279	0.0203	+
9/3	799	0.0747	0.0638	+
8/4	1398	0.1308	0.1353	−
7/5	2033	0.1902	0.2041	
6/6	2360	0.2208	0.2244	
5/7	1821	0.1703	0.1813	
4/8	1198	0.1121	0.1068	+
3/9	521	0.0487	0.0448	+
2/10	160	0.0150	0.0127	+
1/11	29	0.0027	0.0022	+
0/12	6	0.0006	0.0002	+

maturity. In a closed system the controlling agent is some hormonal factors. It will be recalled that the mechanism is statistical, i.e. only affecting the probability of having a boy. At first sight, Geodakyan's model appears to be fairly reasonable, but it, of course, needs careful experimental check.

Caution: the Problem Reduced to a Linear One

The problem has reduced to a linear problem. Mathematically, this implies that all the major difficulties have been left behind, it only remains to make use of a thoroughly developed mathematical tool and the problem at hand has been tackled both from the theoretical and computational points of view.

There is a "but" here. If the problem is practical in nature and a linear model describes some real object, we are still not safeguarded against some surprises.

To begin with, consider a simple system of two algebraic equations in two unknowns, contained in any high-school algebra course:

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1, \\ a_{21}x_1 + a_{22}x_2 &= b_2. \end{aligned}\right\} \qquad (1)$$

Its solution is

$$x_1 = \frac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{12}a_{21}}, \quad x_2 = \frac{b_2a_{11} - b_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}}. \qquad (2)$$

The algebraic theory of systems of linear equations assumes that coefficients a_{ij} and b_i are known accurately, an assumption as acceptable for mathematics as unacceptable for applications. Really, when (1) is a mathematical model of some physical object, the coefficients have some specific sense. They are determined by measurement or computation and frequently rather approximately.

How do uncertainties of the coefficients affect the solution of system (1)? Computer people have long observed that small deviations of the coefficients of (1) can sometimes have disastrous effects. A classical example is the system

$$\left.\begin{aligned} x_1 + 10x_2 &= 11, \\ 10x_1 + 101x_2 &= 111. \end{aligned}\right\}$$

Its solution is $x_1 \quad 1$, $x_2 \quad 1$, and the solution of the system

$$\left.\begin{aligned} x_1 + 10x_2 &= 11.1, \\ 10x_1 + 101x_2 &= 111 \end{aligned}\right\}$$

is $x_1 = 11.1$, $x_2 = 0$.

Such systems are referred to as ill-conditioned, and the first recommendations of mathematicians were as follows: try to find it in due time and avoid it in an application. But it turned out that physical and technical problems frequently come down to an ill-conditioned system, and if we overlook the fact we may appear in the position of that colonel who was the only one to march in step, while the entire regiment was out of step. This revived interest in ill-conditioned systems, and they were brought under the umbrella of so-called incorrect problems.*

* The concept of correctness for boundary problem formulation for differential equations was introduced by the French mathematician Hadamard early in the 1930s and he furnished an example of a situation where a small variation of initial data resulted in an arbitrarily large changes in the solution. Later on the importance of the concept for physical systems was understood, one of the major workers in the field being the Soviet mathematician Petrovsky.

Now the theory of linear algebraic equations must begin with a definition of what is to be understood under the solution to a system of equations. Next, in the usual fashion, we have to examine the system for existence and uniqueness of the solution and find ways of constructing it, and surely see to it that the fresh solution does not respond that violently to minute changes in the initial data, in other words is stable.

Now we would like to discuss systems of linear algebraic equations (1) where a_{ij} and b_i are random variables.

Before we proceed to examine such systems, it is worth noting the general and common character of the model we are going to discuss. Clearly, most problems of mathematical modelling involve examination and solution of finite, differential, integral and more complex equations. In real problems the parameters or coefficients in equations are found experimentally or preset. To work out the problem as a rule needs discretization of respective equations, e.g. transition from differential to difference equations. The discretization is necessary so that digital computers might be used. In the simplest, also commonest, case after the equations have been discretized and simplified the result will be a system of linear algebraic equations.

Recall that we treat the coefficients as random variables. Let us now show by several examples that this is quite reasonable a decision in many applications.

Consider the series oscillatory circuit consisting essentially of a resistor R, capacitor C and inductor L (Fig. 7). If the input voltage (real)

is U, then the components I_1 and I_2 of the complex instantaneous current $I = I_1 + jI_2$ are known to be found from the system

$$\left.\begin{aligned} RI_1 - \left(\omega L - \frac{1}{\omega C}\right) I_2 = U, \\ \left(\omega L - \frac{1}{\omega C}\right) I_1 + RI_2 = 0. \end{aligned}\right\} \tag{3}$$

Now we can, of course, substitute the coefficients of this system into (2) and thus obtain

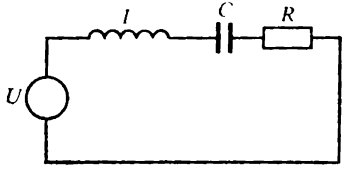


Fig. 7

the solutions for I_1 and I_2. But where do R, L and C come from?

If this is a sample of a batch, then the exact values of R, L and C are unknown. What is more, no experiment will give these values absolutely accurately, since, as it was mentioned above, any measurements are always conducted with a limited accuracy, a principal premise of the theory of measurement. Uncertainties of R, L and C lead to uncertainties in determining I, and solutions to (3) are thus approximate. As is assumed in the theory of measurement, inaccuracies in measurements are treated as random variables, and hence solution I will be a random variable too.

A further example is taken from economic planning. Suppose we need to coordinate the output of three groups of factories including their interrelations and relations with their suppliers and customers. Denote the final products by y_1, y_2 and y_3, and total outputs by x_1, x_2 and x_3. If a_{ij} is the rated consumption of products of ith group of factories to produce one tonne of products in jth group of factories, then the total outputs and final products are related by

$$\left.\begin{aligned} (1-a_{11})\,x_1 - a_{12}x_2 - a_{13}x_3 &= y_1, \\ -a_{12}x_1 + (1-a_{22})\,x_2 - a_{23}x_3 &= y_2, \\ -a_{13}x_1 - a_{23}x_2 + (1-a_{33})x_3 &= y_3, \end{aligned}\right\} \qquad (4)$$

where $i, j = 1, 2, 3$.

The rated consumptions a_{ij} are averaged quantities here and they cannot be specified accurately.

The generalization of the problem to n groups of factories is quite obvious: instead of (4) we will have a similar system of n equations.

We could multiply such examples no end. In other economic, sociological, and even technological, problems that boil down to a system of linear algebraic equations, coefficients a_{ij} at times cannot be derived, and then they are specified by expert estimates. Being apparently subjective, these estimates cannot be viewed as specified exactly.

Thus, if coefficients are obtained by experiment or calculation, carried out with limited accuracy, they can reasonably be thought of as realizations of some random variables. System (1) will then contain six random variables a_{11}, a_{12}, a_{21}, a_{22},

b_1, and b_2. The solution to the system is naturally a random vector. To be more specific: the solution to the system of n linear algebraic equations with random coefficients is the random vector $x = (x_1, x_2, \ldots, x_n)$ such that all the random variables $\left[\sum_{k=1}^{n} a_{ik}x_k\right] - b_i$, $i = 1, \ldots, n$ are equal to zero with probability 1.

Since the law of the joint distribution of the coefficients in a given system is considered known, then, formally speaking, we can derive the n-dimensional law of the joint distribution of the components of the solution-vector $x_1, x_2, \ldots, x_n$, and hence work out the distribution for each of x_i. The calculations are extremely tedious, however. For example, a linear system of the tenth order contains 110 coefficients, and to compute the distribution of each of the components requires taking a 110-tuple integral. Any volunteers? But the main thing here is that it would be more difficult to use this form of solution. In fact, what a physicist would make of the current in an oscillatory circuit specified by a joint distribution of the real and imaginary parts of the current? Fortunately, in engineering applications the main role is played not by distribution itself, but by some numerical characteristics: mean, variance, the most probable value, spread, and so on. Note that such a situation is typical of many applications of probability and statistics.

In what follows we will characterize the random solution-vector of an algebraic system with random coefficients by its mathematical expecta-

tion, i.e. by a vector obtained from a given random vector by replacing all its components by their respective expectations. It would appear that the problem became simpler, but now new difficulties emerged, since the solution of a system with random coefficients may have no expectation at all. The following examples show that such reservations are not ungrounded. Recall Ohm's law

$$I = U\,(1/R), \tag{5}$$

where I is the current, U is the voltage, and R is the resistance. This relationship is a simple linear algebraic equation, where $1/R$ is a coefficient.

Consider the mass production of radio sets. In the same place in the circuit "identical" resistors are installed with a rated resistance 1000 ohms. The voltage U is here fixed at 100 V, say. If the resistor's value is exactly as rated, then the current through the circuit will be $100/1000 = 0.1$ A. As you know, however, resistors, just like any products, are manufactured with errors. Therefore, commercial resistors are generally labelled, say, 1 kohm $\pm$ 5%. Since the time of Gauss experimental errors, as well as manufacture errors, are assumed to be described by the normal distribution. If R is a random variable, then the current flowing through the circuit will be a random variable as well.

What average current will be in the circuit? The word "average" here is to be understood as the average over the ensemble of manufactured circuits. Since the random variable R is in the denominator, this average (expectation of cur-

rent) will not be equal to the value 0.1 A, which is obtained by substituting the rated value into the denominator. What is more, in the accepted model the desired expectation simply does not exist.

To explain, we will have to write an integral, but, as it was said earlier, if you believe me and do not want to bother yourself with computations, just skip them.

For a normally distributed R we have

$$\mathbf{M}I = \frac{U}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{R} e^{-\frac{(R-R_0)^2}{2\sigma^2}} dR, \qquad (6)$$

where $R_0 = \mathbf{M}R$, which in our example is equal to 1000 ohms. But this simple integral diverges due to a singularity of the first order at $R = 0$, and there is no $\mathbf{M}I$.

Returning to the system (1) and its solution (2), we will consider the simple situation where $a_{11} = y$ is a normal random variable with expectation a and variance σ^2, the remaining a_{ij} and b_i being some constants. The expectation x_1 will then be

$$\mathbf{M}x_1 = \int_{-\infty}^{\infty} \frac{b_1a_{22} - b_2a_{12}}{ya_{22} - a_{12}a_{21}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-a)^2}{2\sigma^2}} dy. \qquad (7)$$

Since the denominator of the first term of the integrand vanishes at $y = a_{12}a_{21}/a_{22}$ (a first-order zero), and the second term never vanishes, the integral (7), similar to (6), is bound to diverge*.

* Assume that the numerator is nonzero, otherwise we will have to consider $\mathbf{M}x_2$.

And so, in a situation where one of the coefficients is random and normally distributed, the solution components have no expectation.

Now, clearly, if not only a_{11} but also all the coefficients are random and normal, then, generally speaking, the components have neither expectation nor variance, because there are always some combinations of coefficients available at which the denominator vanishes and the integrals will diverge. In the above example of the oscillatory circuit reducing to (3), manufactured resistors, inductors and capacitors always show some spread in R, L, and C, respectively. According to the above-mentioned assumptions of error theory, these errors must be described by a normal distribution. The above reasoning suggests that the solution has no expectation and variance, i.e. the average current in the circuit is nonexistent, and its variance, and hence power, are infinite. No electrical engineer, I think, will agree to it.

Random variables having no expectation and variance, as a rule, are of no interest in real problems, and therefore in a mathematical model we should beforehand exclude situations where the solution components have no expectation and variance.

I first got exposed to such issues in connection with some problems in geophysics. Strange as it was, it turned out that when the coefficients were described by some most common stochastic models, their simplest numerical characteristics, such as expectation and variance, were generally nonexistent.

Quantities with which we dealt were, however,

quite real: velocities of propagation of elastic waves in the Earth's crust, depth of horizons, and so on. Therefore, the absence of expectation is, obviously, at variance with common sense and practice. The only conclusion is: all the worse for the model, since it is responsible for such discrepancies, the model is inadequate, it should be changed, replaced by another one. It is, of course, useful, and even necessary, to try and understand the reasons why the model does not correspond to the real situation.

It turned out that the disagreement is due to taking into account the wings of distributions. In systems of linear equations with random coefficients we abandoned distributions similar to the normal or exponential ones, with wings going to infinity, and turned to distributions concentrated in finite intervals, or, in the language of mathematics, finite distributions. Examples of finite distributions are the uniform distribution and the so-called truncated normal distribution, i.e. a normal distribution in which all the values are discarded that lie beyond some interval, or as they say, the wings of the distribution are severed.

The formal consideration of such finite distribution in general has no principal limitations, if only of physiological character: everybody believes that the normal distribution is versatile. But any versatility, and here too, has its boundaries. And if the normal, exponential or any other distribution does not suit you because of the wings, do not hesitate to sever the wings and use a finite distribution as your model.

Solution: Formula or Number?

To be sure, you know the answer: it depends... A formula, if it is sufficiently simple and graphic, makes it possible to see the qualitative picture: the variation of the solution as a function of the parameters, the behaviour of the solution at very large or very small values of the variables, and so on. This information is at times necessary not only for the theoretician, but also for the practical engineer, experimenter, economist. But sooner or later the solution must be expressed in numbers. And then, if the formal solution of the problem is complicated or unavailable, a problem emerges of calculability of the solution: to devise a method of calculating approximate solutions or a calculational technique that will enable ever increasing accuracy to be achieved. This is the situation that emerges in the problem discussed in the previous section. Computation of the expectation and variance boils down to multiple integrals. But to write down expressions and to obtain numerical values are different problems. Numerical quadratures of multiple integrals at times give rise to enormous computational difficulties.

The commonest methods of deriving the numerical characteristics of components of the solution of a system are either statistical modelling (Monte-Carlo method) or the expansion of the solution into the Taylor series in the neighbourhood of the mathematical expectation of the coefficients of the system.

In the Monte-Carlo method we should, according to Cramer's rule, write the formal solution

of the system, i.e. equations (2) of the previous section, and then, in accordance with distributions of the coefficients, select from the table of random numbers their values, compute the determinants that enter Cramer's formulas, so that to obtain realizations of the components of the solution vector. Next, after having accumulated enough realizations, i.e. a sample of a sufficient size, we should take the empirical average, i.e. the average of arithmetic sample.

If the order of the system is not too low, however, then the Monte-Carlo method involves tedious computations and the approximations obtained are relatively inaccurate.

The second method harbours a couple of sunken reefs. Return to the oscillatory circuit of the previous section, i.e. to system (3) there.

To be more specific, suppose that R, L and U are constant quantities that take on the values R 1 ohm, $L = 10^{-3}$ Hz, $U = 5$ V, whereas C is a random variable uniformly distributed over the interval $\gamma = (10^{-9} - 0.01 \times 10^{-9}, 10^{-9} + 0.01 \times \times 10^{-9})$, i.e. in the range 10^{-9} F $\pm$ 1%. The modulus of the complex current $| I |$ will then be a definite function of C, and so we will be able to find it readily. To determine the expectation of the current modulus at the resonance frequency $\omega = 1/\sqrt{L \cdot \mathbf{M}C}$ we will have to take the integral of the current modulus over the interval γ (by dividing the integral by the interval length), and we will obtain $\mathbf{M}I = 1.5$ A.

If, as recommended in practical texts, we expand $I = I(C)$ into the Taylor series, keeping the first two terms, and now compute the expectation, we will get 5 A. Thus a 1 per cent error

in capacitance yields an uncertainty in MI that is as high as 330 per cent.

Note that the use of the first two terms of the Taylor expansion to find the expectation of the solution is equivalent to replacing the coefficients of the system from the previous section by their expectations, so that the new system of equations is solved using deterministic coefficients. This path in general seems to be attractive: it appears that by replacing the coefficients by their mathematical expectations and solving the resultant system of algebraic equations, we can obtain a satisfactory estimate of the expectation of the solution.

Consider a further example illustrating the extreme roughness, and hence unsuitability, of such an approach without preliminary estimation of the possible error.

We would like to find the extremum of the parabola

$$y = ax^2 - 2bx, \qquad (*)$$

where a and b are independent random variables, distributed uniformly within the intervals $(10^{-3}, 1)$ and $(5, 7)$, respectively.

The stationary point x_0 and extremum y_0 are here random variables, too.

$$x_0 = \frac{b}{a}, \qquad y_0 = -\frac{b^2}{a}. \qquad (**)$$

If now we take into account the distribution of these random variables, we obtain for their expectations

$$Mx_0 = 87.3, \quad My_0 = -251.1.$$

If in (*) and (**) we replace a and b by their expectations $\mathbf{M}a = (1 - 10^{-3})/2$, $\mathbf{M}b = 6$ and calculate the coordinate $\hat{x}_0$ of the stationary point x_0 and the extremum y_0, then

$$\hat{x}_0 \approx 12, \quad \hat{y}_0 = -72.$$

So the error turns out to be unacceptably large, and it really may be arbitrarily large with a "bad" distribution of a. For example, if a is distributed uniformly within the interval $(-0.5, 1)$, then $\mathbf{M}a = 0.25$ and estimates $\hat{x}_0$ and $\hat{y}_0$ assume finite values, whereas in reality there are no expectations $\mathbf{M}x_0$, $\mathbf{M}y_0$, because $\int_{-0.5}^{1} \frac{da}{a}$ diverges.

The above examples and other problems of practical interest show that these ways of finding the expectation of a system of linear algebraic equations with random coefficients may give very rough estimates and mistakes.

This should be especially taken into account, when the coefficients are interdependent. Here is a paradoxical example. Let the matrix of the coefficients of system (1) of the previous section be

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix},$$

where α is a random variable uniformly distributed over the interval $[0, 2\pi]$. At any realization of α we have $\det A = 1$, and accordingly the solution of (1) where the coefficients are replaced by the elements of matrix A exists for any vector

b, and $\mathbf{M}x_1 = \mathbf{M}x_2 = 0$. At the same time all of the elements of A have zero expectations, and so, if we substitute expectations for the coefficients of the system in question, then the resultant equations will have no solution at all, as well as the expectation.

Note that this example is by no means artificial, because A is the matrix of the turn of an orthogonal frame of reference through an angle α.

It can be expected that the expectation of the system's solution will be found to higher accuracy, if we have more terms in the Taylor expansion. This is the case in practice. The finding of the next terms of the expansion is, however, fraught with enormous computational difficulties, since the number of terms grows very fast (remember that we here deal with a series for functions in many unknowns), and their form varies from problem to problem. Besides, simply to increase the number of terms in the expansion without taking care of the accuracy of computations of the expectation may lead to an absurdity.

For example, if in seeking the expectation of the current modulus in an oscillatory circuit we keep three terms, not two as before, the result obtained ($\mathbf{M}I = -83.3$) will make no sense, because the modulus of I, and hence its expectation, are positive.

At the same time the Taylor representation of the solution of (1) has one very important feature. Unlike the solution in the Cramer formula, which is a fractional rational function of many variables, i.e. coefficients, the truncated Taylor

series is a polynomial in those variables. Therefore, if the truncated power series represents the solution sufficiently well, then the very computation of the expectation (i.e. taking the integrals of an algebraic polynomial) is smooth sailing. It would make sense, therefore, to reject not the representation of the solution by a power series, but only the use of a Taylor series. We would now like to require that the new series would, first, converge sufficiently fast, and second, that its terms would follow fairly simply, and if possible uniformly, from the coefficients of a given system of equations; and third, that the accuracy of the remaining term would be calculable. All of these requirements can be met, if we use iteration methods in seeking an approximate solution.

I took the case of solving a system of linear algebraic equations with random coefficients to illustrate the critical manner in which the solution is conditioned by the model selected, in particular in the above problem, by the distribution type. If the probabilistic model of random coefficients is taken to be a normal distribution, or its equivalent, with wings going to infinity, there will be no reasonable solution. If we "sever" the wings and think of the models of the coefficients as finite distributions, we will obtain reasonable results. Therefore, before setting out to solve the problem, we must pay attention to the selection of the mathematical model.

Identification of Criminals—Bertillon System

The problem of identification of persons is a very difficult one. One hundred years ago criminal police in many countries of Europe used to compare photographs (full-face and profile) and some descriptions, but just think of sifting tens of thousands of pictures to select similar ones. Therefore, the problem of identification of criminals was a very burning issue.

In 1879, in the Paris police prefecture, appeared a new clerk, Alphonse Bertillon. His task was to fill in cards with descriptions of criminals. The notes were fairly indefinite: tall, medium height, or short, scarred face or not, or just "no special features".

Bertillon was born into a family fond of natural sciences—his father was a respected physician, statistician and vice-president of the Paris bureau of vital statistics. He read Darwin, Pasteur, Dalton, Gay-Lussac, and heard of Adolphe Quételet, a Belgian mathematician and statistician, who is not only remembered for his mathematical studies, but also for his proof that the human body measurements are governed by certain laws. Further I selectively quote from Thorwald's book *One Hundred Years of Criminalistics*, which is a fascinating work of science where detective stories illustrate scientific advance of criminalists, and therefore the book is a riveting reading.

"And so in June of 1879, when Bertillon, exhausted with the Paris heat, was sitting and filling in the three or four-thousandth card till

he was blue in the face, he was suddenly hit by an idea, which was born, as he later confessed, by the awareness of absolute senselessness of his work and at the same time by his childhood memories. Why, he asked himself, are so time, money and efforts wasted to identify criminals? Why stick to old, crude and imperfect methods, when natural sciences provided possibilities unfailingly to distinguish one man from another by the size of the body?

"Bertillon evoked surprise and derision of other clerks, when at the end of July he set out to compare photographs of prisoners. He compared the shape of ears and noses. Bertillon's request to allow him to measure the checked-in prisoners raised uproaring laughter. But much to general joy the permission was granted. With gloomy and bitter zeal he had in several weeks taken measurements of quite a number of prisoners. In measuring their heights, lengths and volumes of a head, the length of hands, fingers, and feet, he saw that sizes of individual parts of the body of various persons may coincide, but the sizes of four or five parts would never be the same.

"The stuffy heat of August caused fits of migraine and nasal bleedings, but Bertillon, however useless and purposeless it would seem, was captured by the 'power of the idea' In mid-August he wrote a report explaining how it was possible to identify criminals without fail. He addressed the report to the prefect of Paris police, but got no answer.

"Bertillon continued his work. Each morning before work he visited the prison La Santé, There

he was also made fun of, although allowed to take his measurements. When on October 1 he was promoted, he submitted to the prefect a second report, in which, referring to the Quételet law, he noted that the sizes of bones of an adult remain unchanged throughout his or her life. He maintained that if the probability of a coincidence of the heights of people is 4 1, the height plus another measurement, for example, the length of the body to the waist, reduce the probability down to 16 1. And if 11 measurements are made and fixed in the card of a criminal, then the estimated probability of chancing upon another criminal with the same statistics will be 4,191,304 1. And with fourteen measurements the chance will reduce to 286,435,465 1. The set of members that can be measured is very large: in addition to the height of a man, measured can be the length and width of his head, the length of fingers, forearm, feet, and so on. He wrote: 'All the available identification techniques are superficial, unreliable, imperfect and give rise to mistakes.' But his technique makes one absolutely confident and excludes mistakes. Furthermore, Bertillon worked out a system of registration of cards with measurement results, which made it possible in a matter of minutes to establish whether or not the data on a criminal were available in the file."

Thus Bertillon suggested to make use of a set of anthropological data to identify criminals.

To be sure, much time and effort were required to overcome stagnation and mistrust. But success and recognition came to Bertillon, as usual, due to a fortunate coincidence, when his regis-

tration system and the enormous work enabled several major criminals to be identified.

Bertillon's system, or bertillonage, consisted in measuring the height, the spread of arms, the width of the chest, the length and width of the head, the length of the left foot, the length of the third finger of the left hand, the left ear, and so on.

Now let us take a closer look at bertillonage. This is essentially a mathematical model of a man in the form of a set of numbers $(x_1, x_2, \ldots, x_n)$, that is in the form of a point in the n-dimensional space or n-dimensional vector.

Bertillon relied on the calculated probabilities of meeting two persons with the same values of sizes. The statement "there are no two men on earth such that the sizes of individual parts of their bodies coincided and the probability of meeting two people with absolutely the same height is estimated to be 1 : 4" in the book by Thorwald is ascribed to Quételet. Thorwald also maintains that the father and grandfather of Bertillon (the latter was a mathematician and natural scientist) have tested Quételet's statement.

It seems to me that in those computations at least two mistakes are made. First, the probability for the heights of two randomly selected people to coincide is not 1 : 4, but three to four times smaller. Second, in the above calculation multiplied together are the probabilities of coincidences of the sizes of the chosen parts of the body, that is, statistical independence is assumed of, say, the width and length of the head or the height and the spread of arms. But here we have no statistical independence, since these quanti-

ties are strongly correlated and therefore we are surprised when we encounter a tall man with a small head or a short man with long arms.

The mistakes in the calculation make up for one another to a certain degree, and so the probabilities of having two people with the same sizes of all the 11 quantities are really exceedingly small, thereby making bertillonage so successful.

For a time bertillonage became universally accepted but its widespread uses were hampered by a number of circumstances, the major one being the complexity of realization. To take measurements it is necessary that the person being measured cooperated in the operation: he must sit still and offer his head, hand, foot, and so forth. The person responsible for the measurements must act accurately and carefully. The cultural level of policemen and jail people being rather low, the results of measurements they carried out could not be trusted. And so bertillonage, although it had gained a measure of recognition in some countries, failed to become a common method.

Such a situation obtains fairly often: an impeccable theoretical work does not find its way into practice due to the complexities of its realization. Bertillon failed to simplify his system so that taking measurements would not require high skill of personnel. Therefore, in a time bertillonage was replaced by dactyloscopy, a comparison of finger prints. Its history is rather instructive, but it concerns other topic—image recognition—which is not to be discussed here.

Identification of Objects

Crude oil, as produced, contains a sizable amount of water. The water extracted from an oil-bearing horizon, called oil-formation water, causes much trouble: it is strongly mineralized, contains up to thousands of milligrams of salt per litre.

It is well known that oil is lighter than water, and so at oil fields the bulk of water is separated from oil by settling in reservoirs. After the settling the mixture is separated: the upper layer of oil is pumped to oil pipe lines, and the lower part, water, is pumped back underground, through injection wells. This method, however, does not enable one to get rid of water and salts completely, and so some of this harmful mineralized water (up to several per cent of oil volume) is transported together with oil to oil refineries. This salt is responsible for fast corrosion of metallic parts, and if salt is not extracted from oil before processing, equipment will always fail, and oil products, especially residual oil, will be of inferior quality. Therefore, refineries have long been equipped with electric desalting plants (EDP), so that oil is desalted before primary processing.

The principle behind EDP is simple. It is common knowledge that water will not dissolve in oil, and so small droplets of water, from microns to fractional millimetres in size, are suspended in oil. It is these droplets that contain the salt. Under gravity the droplets settle down to the bottom of pipe-lines or reservoirs, thus forming bottom layers of water, which can be removed.

The precipitation rate varies with the square of the droplet size. Accordingly, if the water droplets are made larger, they will settle down quickly, thus segregating from oil. To this end, oil is "washed out": a large quantity of weakly mineralized water is added to it so that small droplets of strongly mineralized formation water merged with droplets of washing water, formed droplets of larger size, which now will settle faster.

Clearly, a drop formed by merging of a fresh and saline droplets will have lower salinity than the initial drop. To intensify the merging and settling of droplets in EDP the oil-water mixture is fed through a high electric field. Economic pressure required that electric desalting efficiency be enhanced as much as possible. And so the optimization problem here became extremely important.

Being extremely involved, the process could not be described by a model covering all the spectrum of physical and chemical processes, and so models are generally constructed on the basis of evidence derived when an object of interest to us functions normally. The model must be suitable for the object to be controlled at optimal or near-optimal regime. The model must thus describe the process adequately, and be its counterpart in the control system.

We are here confronted with the problem of choosing among the many models possible, and comparing the model and the object to test if the model selected can be used as the counterpart, i.e. the problem of identification of the model with the object, the original. "Identification" here should not, of course, be understood literally,

as in the case of the identification of criminals. No mathematical model can describe an object perfectly. Hence the model selection problem. The object itself is thought of as a "black box", i.e. only some of its inputs and outputs are taken

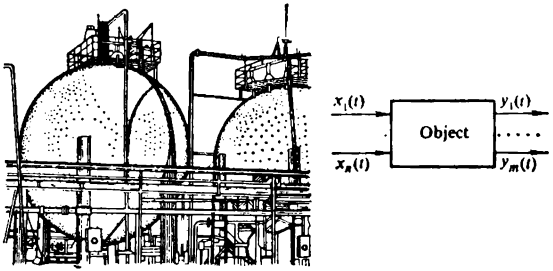


Fig. 8

into consideration, ignoring the information about the state of the object itself. Put another way, the model does not include information about processes occurring within the plant, as well as other inputs and outputs, which are taken to be either fixed, or small.

Figure 8 shows a real EDP and its model in the form of "black box" with some inputs and outputs.

The diagram of Fig. 8 can be simplified by thinking of $x(t)$ and $y(t)$ as some vector-functions with the coordinates $x_1(t)$, $x_2(t)$, ., ., $x_n(t)$ and $y_1(t)$, $y_2(t)$, $y_m(t)$, respectively. The resultant diagram is given in Fig. 9.

Modelling of dynamic systems in mechanics, automatics, radio engineering makes especially extensive use of differential equations.

Earlier the identification problem was solved as finding the coefficients in a similar equation from experimental data accumulated in commercial operation.

However, in identifying complex systems we cannot always consider that the outputs and inputs are related by a linear differential equation. Selection of model type is one of the basic and

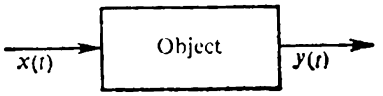


Fig. 9

most difficult problems in modelling. A real-life object shown schematically in Figs. 8 and 9 effects transformation of the function $x(t)$ into function $y(t)$, and the model somehow or other reflects this transformation.

But back to the electric desalination of petroleum. The lower the concentration of salts at the EDP input, the lower it is at the output—here is the principal idea behind the procedure, and this is quite reasonable. It follows that to reduce the salt concentration at the EDP output several times we will have to reduce the input concentration several times, i.e. it is necessary even at the oil fields to have heavy-duty industrial EDPs. This can be done, but it would require enormous investment. Perhaps we can do with simpler means—to optimize EDPs at refineries, so providing high-quality desalting?

If such a solution of desalting problem is possible, it will be worth having because it will be cost-effective.

Before we set out to consider the optimization, we will take a quick look at the principle of EDP operation. Denote by x the input salt concentration and by y the output concentration.

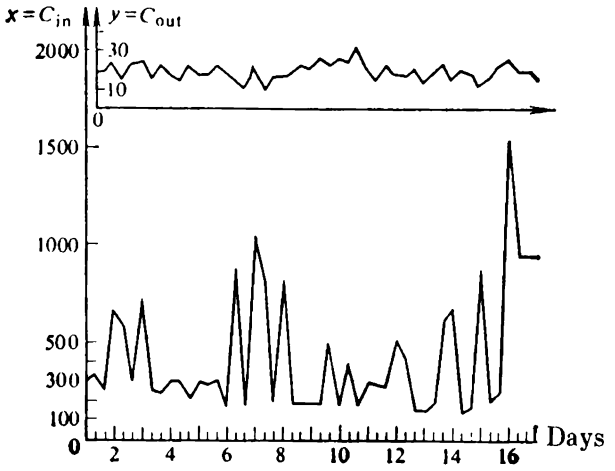


Fig. 10

Mathematically, EDP transforms x into y at each moment of time, but the function describing the transformation is unknown.

Figure 10 gives the variation of x and y with time at real EDPs (about 15 all in all). The variation is highly irregular, the lower curve's shape is conditioned by the irregular character of operation of crude oil dehydration plants at oil fields, and by the mixing in pipe-line of oils derived from various horizons or boreholes, and

oil fields, and so forth. It is thus impossible to ensure constant concentration of salts at the input or, at least, its smooth variation.

Compare the two curves. Could we maintain that there is some functional dependence between

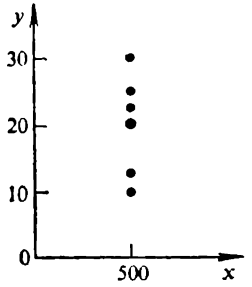


Fig. 11

the two curves? No, we could not. We will thus have to get some reasonable model and process the data available.

To test the initial hypotheses concerning the input and output concentrations we will proceed as follows. Let us fix some input concentration, say, 500 mg/litre and collect all the output data for this value. If the hypotheses were exactly true, then the output data for 500 mg/litre at the input would be the same. But in reality this is not so, and in Fig. 11 we can see the range of their variation.

This is not surprising and we can really account for this spread: any process is affected by a large number of interferences, especially in such a "noisy" process as the EDP one, where a wide

variety of unaccounted parameters of electric, hydrodynamic and processing characters are available. The interferences here may be got rid of partially by averaging. Since the expectation of y is taken for a fixed value $x = 500$ mg/litre, this

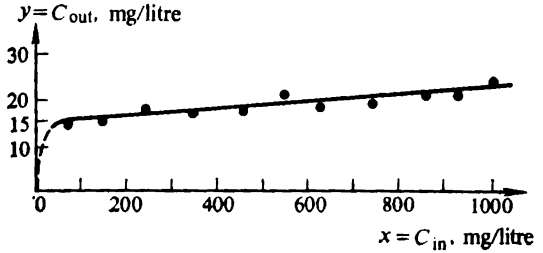


Fig. 12

expectation is a conditional one. This is not the whole of the story, because computation of the expectation requires the probability distribution of a random variable in question, which is unknown: we have only a set of observed values, i.e. a sample. As is usual in such a situation, we here take the arithmetic mean and thus obtain the conditional empirical mean, which in Fig. 11 corresponds to a large point. The further procedure is clear. To begin with, we plot the corresponding points y for various values of x, and so we obtain in the plane (x, y) a cloud of values. Next for each value of x, i.e. on each vertical, we average the data available and obtain a set of large points (Fig. 12). These large empirical points are analogues of the conditional expectation of the output for a given input. It is at this

point that we should seek a dependence between y and x, or in our specific problem, the relationship between the salt concentrations at the input and output.

In our case the data at this point turned out to be both simple and unclear. It might seem that the points lie on a straight line. To be sure, it should be taken as a mathematical model of the relationship between the output and input for the plant under consideration. Some spread of the empirical means derived is quite understandable: the data are obtained for a normal operation of the plant, their body is limited and the averaging carried out cannot fully protect us from errors. But notice the gentle slope of the line, it is nearly parallel to the x-axis, and hence the plant appears to be but weakly sensitive to variation of the salt concentration at the input: changing the input concentration from 1000 mg/ /litre to 100 mg/litre, i.e. tenfold, produces but barely perceptible changes in the output concentration from 24 mg/litre to 16 mg/litre, i.e. by a factor of 1.5, with 1 mg/litre lying within the accuracy of the measurement? What are we to make of such a dependence? We are led to conclude that the plants are essentially insensitive to the input salt concentration, they function inadequately, and even at small input concentrations (about 100 mg/litre) they do not ensure a noticeable reduction of the output concentration. Thus, the initial assumption that for the output salt concentration to be lower the input concentration must be lower is erroneous.

We now must construct the model of EDP operation to reflect the relationship between input

and output salt concentrations. According to the hypothesis of the random structure of the interferences the conditional expectation of the output, given the input, gives a linear equation. Straight lines give an adequate description of experimental evidence. But if we take a closer look at the curves obtained we will find some contradiction: we may not extrapolate the line to the region from zero to 100 mg/litre, since the plant does not enable output concentration of, say, 10 mg/litre to be obtained when the oil at the input has a salt concentration of 5 mg/litre. Therefore, in the interval $0 \leqslant x \leqslant 100$ the model as constructed, i.e. the straight line does not work, and another one is necessary. Clearly, the function must pass through the origin of the coordinate system: if oil at the input of an EDP is free of salts, then at the output there will be no salts either. Since for the interval $0 \leqslant x \leqslant 100$ there is no experimental evidence, a new model must be constructed based on some qualitative physical considerations. We took it to be an increasing exponent for this interval, as shown by the dash line in Fig. 12.

I subjected you to the tedium of a detailed examination of a specific example so that you could trace the sequence of steps in constructing a mathematical model to identify a complex object, and see the reason why such a model may be needed and what benefits could be derived by carefully analyzing the results of object identification.

Further analysis of the results obtained here indicated that no optimization can be effected at the expense of changing the characteristics

and parameters of an existing EDP Consequently, it was here necessary to give a careful treatment to the physics and even physical chemistry of the desalting process. As a result, it turned out that for coalescence and separation of droplets to occur more efficiently it was necessary drastically to increase the time the emulsion spends in an electric field. To implement this recommendation, a special-purpose device, the electrocoalescentor, providing the required residence time was designed and manufactured. In addition, a deeper understanding of physical and chemical laws governing the process enabled a recommendation to be worked out to change the amounts of flushing water, amounts and points of feeding demulsifier, to correct some of the processing parameters and ultimately to reduce the residual salt content threefold or fourfold, thus achieving substantial savings of funds, reagents and fresh water.

Regression

Returning to the set of input-output data, we obtain a similar set if the point coordinates will be length x and diameter y of trees in a grove, or length x and width y of a criminal's head in the Bertillon classification. So now we may for the moment ignore the real content of the quantities x and y and formulate the general problem.

Let there be two random variables ξ and η, related by some dependence: it seems that the larger the value x that ξ takes, the larger the value y taken by η, or y decreases with increasing

x, or the dependence between x and y is given by a quadratic or other function. The situation is uncertain: there is no clear, definite, deterministic dependence. This dependence is statistical, it only

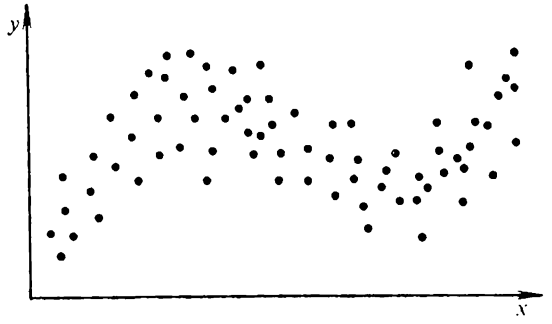


Fig. 13

shows up on average: if we draw the set of data corresponding to realizations of the random variables, (ξ, η), i.e. to the observed pairs of values, points (x_i, y_i), $i = 1, \ldots, n$, then they lie on some curve. Figure 13 illustrates such a situation where, unlike Fig. 12, the points concentrate along a curve displaying a distinct maximum and minimum.

Theoretically, this curve can be found fairly simply, if the pair (ξ, η) is given by a joint probability distribution: it is then that we should plot the curve for the conditional expectation of the random variable η given that the random variable ξ assumes the value x:

$$y = \mathbf{M}(\eta | \xi = x) = \varphi(x).$$

This function is the desired relationship "on average" between η and ξ. The equation $y = \varphi(x)$ is called the *regression equation*, or rather the regression equation for η on ξ, because we can also consider the equation $x = \mathbf{M}(\xi \mid \eta = y) \quad \psi(y)$, i.e. the regression equation for ξ on η, where the curves for $y = \varphi(x)$ and $x = \psi(y)$, generally speaking, do not coincide.

The word "regression" was introduced into statistics by Sir Francis Galton, one of the originators of the science of mathematical statistics. Correlating the heights of children and their parents, he found that the dependence is but slight, much less than expected. Galton attributed it to inheritance from earlier ancestors, not only parents: according to his assumption, i.e. his mathematical model, the height is conditioned half by the parents, a quarter by the grand-parents, and one-eight, by great-grand-parents, and so on. There is no saying here whether Galton is right, but he paid attention to the backward motion in the family tree, and called the phenomenon regression, i.e. motion backwards, unlike progression, the motion forwards. The word "regression" was destined to occupy a prominent place in statistics, although, as is so often in any language, including the language of science, another sense is now read into it—it implies a statistical relation between random variables.

In actual practice we nearly never know the exact form of the distribution obeyed by the quantity, and all the more the form of a joint distribution of two or more random variables, and so we do not know the regression equation $y = \varphi(x)$ either. At our disposal there is only some

ensemble of observations, and also a possibility to build models of the regression equation and test them basing on these data. As we have seen above, it was only natural to describe the electrical desalting of oil by a linear function. We can write it as

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

where β_0 and β_1 are the coefficients to be determined from experimental evidence, and ε is the error believed to be random and to have a zero expectation and independent values at various points (x_i, y_i).

If the cloud of data has another form, e.g. as shown in Fig. 13, where it cannot be described by a straight line, then we face a problem of selecting a function for the model of an unknown regression equation.

Building Blocks

Children like constructing castles of blocks, grown-ups use blocks or bricks to erect dwellings and factories, hospitals and barns. Dwellings should not be similar, both in their external appearance and in their "guts": different families need different flats, twin houses are dreary to look at.

Bricks are a good building material, but larger blocks are more economical. But here a problem arises: how is the set of standard blocks to be selected, so that the house would be comfortable, inexpensive and easy to construct.

The main object of calculus is the function, i.e. a dependence of one variable on another, or others, if there are many. Conceivable functions are so many that it would be hopeless to attempt to visualize their diversity. Fortunately, the engineer, biologist or economist does not need this, i.e. he does not need to know exactly

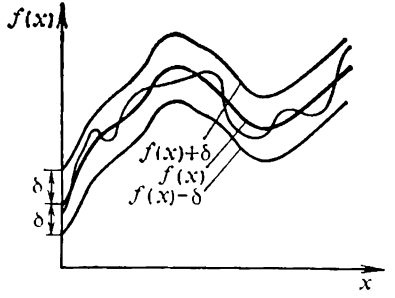


Fig. 14

the behaviour of a function as independent variables vary. In fact, a builder will be content with an accuracy to within a millimetre, a radio engineer normally obtains characteristics of electronic devices to within several per cent, a physician deals with temperature curves of his patients to within one tenth of a degree. Even the trajectory of a spaceship must be known to a finite accuracy.

An accuracy to which a function must be known at each point is a certain number δ, positive of course. It is generally conditioned by the character of a problem, our possibility and desire.

For the situation to be represented dramatically we will on either side of the curve in Fig. 14 draw a line separated from the curve so that its

distance from the line in the direction of the vertical axis will be 2δ. To obtain it in practice we will just shift the curve up and down by δ. Any curve that wholly lies within the confines of the band will be drawn within the same accuracy as the initial one.

A question now arises: if we do not need to know the function with absolute accuracy, and if we can ignore fine details, would it not be possible to replace an arbitrary function by any one close to it, but a simpler one, which is more amenable to examination? You will have now surmised that the answer is positive—why then the above lengthy speculations.

But that is not all that is to it. It turned out that functions that serve as ever closer approximations of a given (any one, though) function can be pieced together from simpler functions, or blocks. To be more specific, we will consider continuous functions. Their plots display no discontinuities, and so are conveniently visualized as threads. There is an infinite variety of continuous functions. For example, there are continuous functions that at no point have a derivative, i.e. no tangent to the curve. It is hard to think of such functions. But for a practician such functions are of no interest, since they cannot be realized in any physical system.

Among the many continuous functions are our old acquaintances, polynomials. A polynomial of degree n in x is generally given by

$$P_n(x) = a_0 + a_1 x + \quad + a_n x^n.$$

It has $n + 1$ coefficients $a_0, a_1, \quad ., a_n$ which are arbitrary real numbers. We can easily plot

the approximate curve of a specific polynomial with numerical coefficients, i.e.

$$P_0(x) = 12 - 4x + 18x^5 - 0.01x^9.$$

To do so, we give various numerical values to the independent variable x in the interval of interest to us, substitute them into the equation and compute the resultant algebraic sum.

In general, polynomials hardly appear so complex functions. What is your estimation of their diversity? Before we go on, think about it.

The simplest among polynomials are the power functions, x, x^2, . . ., x^n, . We should also add 1, a power function of zeroth degree.

Multiplying the power functions by respective coefficients gives

$$P(x) = 5 - 3x - 2x^2 + 5x^3.$$

Any polynomial can thus be pieced together from building blocks, power functions.

Let us now take an arbitrary continuous function in a selected segment $0 \leqslant x \leqslant 1$ and draw a δ-band around it. As we have found earlier, any continuous function whose plot lies wholly within the δ-band is to within δ indistinguishable from it. It appears that among the functions whose curves wholly belong to the δ-band there is a polynomial too.

It is worth noting that the fact is both paradoxical and fundamental: however complex (with angles, sharp variations, etc.) is a continuous function and however small δ, there will always be a polynomial coincident to within δ with the specific continuous function.

My question about the way in which you imagine the diversity of polynomials is quite an insidious one. But you should not be distressed, if you thought that the set of polynomials was simpler than in reality. The above statement is one of the most fundamental theorems of calculus. It is generally known as the Weierstrass theorem about approximation of continuous functions by polynomials.

Karl Weierstrass (1815-1897) is one of the greatest figures in the mathematics of the 19th century. Among the celebrated pleiad of the 19th century mathematicians who reformed the science of mathematics and put it on a more rigorous and advanced foundation the name of Weierstrass shines like a star of the first magnitude.

But back to Weierstrass's theorem. It can also be interpreted in such a way. Let there be a specific, but arbitrary continuous function $f(x)$ and any sequence of decreasing and vanishing numbers, e.g. $\delta_1 = 10^{-1}$, $\delta_2 = 10^{-2}$, . . ., $\delta_n = 10^{-n}$, . . . According to the theorem, for each of these δ_n we can have a polynomial that to within δ_n will be identical to the function $f(x)$ to be approximated. If the polynomials are denoted by $P_1(x)$, $P_2(x)$, . . ., $P_n(x)$, . . ., respectively, then we will obtain a sequence of polynomials ever closer approaching $f(x)$. Since the sequence of δ_n tends to zero with increasing n number, in the limiting case the sequence of polynomials will give the desired function $f(x)$. Thus, the sequence of approximating polynomials describes the behaviour of the initial function $f(x)$.

To sum up, it is the Weierstrass theorem that enables the practician to ged rid of the dismaying diversity of continuous functions and, when necessary, manipulate with polynomials only.

To be sure, the higher the desired accuracy of the approximation (i.e. the lower δ), the higher, in general, will be the degree of the approximating polynomial. But still, polynomials are more convenient to study than arbitrary continuous functions.

But polynomials are far from being the only building blocks from which functions can be constructed to approximate an arbitrary continuous function with a predetermined accuracy.

Turning to Geometry

Geometrical concepts in conventional three-dimensional space are very graphic, and dealing with more general spaces wide use is made of analogies from three-dimensional space: many facts of Euclidean geometry hold true for multidimensional spaces, and those requiring clarifications generally draw on habitual geometric notions. Therefore, the terminology is mostly the same as well.

We will now consider certain of the notions of the theory of multidimensional spaces. Some of them, of course, require exact definitions and proofs. But to illustrate the geometrical meaning of some facts useful for what follows we can do with analogies and speculations, therefore our reasoning will not be that rigorous and will mostly rely on the reader's geometrical intuition.

If you are familiar with the essentials of the theory of multidimensional spaces, including finite-dimensional ones, then you might as well skip the several pages.

You will remember that the vector is a quantity that has magnitude and direction. The angle between vectors can conveniently be given by the scalar product. So if x and y are vectors, then their scalar product (x, y) is the product of their magnitudes multiplied by the cosine of the angle between them. In the theory of multidimensional vector spaces it is convenient to have as the initial concept the above scalar product, which is specified by axioms. We will not here be concerned with them. It is only worth noting here that the square of the vector length is equal to the scalar product of the vector by itself

$$(x, x) = \| x \|^2,$$

where $\| x \|$ is the length, or norm, of the vector. The angle α between x and y is also given by the scalar product

$$\cos \alpha = \frac{(x, y)}{\| x \| \ \| y \|}.$$

If the angle between x and y is right, then their scalar product is zero. Such vectors are called orthogonal, and in elementary geometry, perpendicular.

The set of vectors that can be composed (using the parallelogram rule) and multiplied by numbers form a linear vector space. It can be not only two-dimensional or three-dimensional, like our common space, but may have any number of dimensions. The number of dimensions, or dimensionality, of space is determined by the

largest number of mutually orthogonal vectors that can be arranged in that space. The ensemble of such vectors are generally referred to as the orthogonal basis. It is natural to have them as the coordinate axes, and then each vector can be decomposed into components. If x_1, x_2, . are the projections of x on the unit vectors e_1, e_2, ., then we will have the generalized Pythagorean theorem:

$$\| x \|^2 = \sum_k x_k^2. \qquad (*)$$

If the space is finite-dimensional, i.e. the largest number of mutually orthogonal vectors is finite, then (*) is sufficiently clear. But in an infinite-dimensional linear vector space there is an infinite number of mutually orthogonal vectors, and then in (*) we can assume that the series converges. Such a space is named a Hilbert space after the famous German mathematician David Hilbert (1862-1943), who in 1904-1910 first used geometrical concepts of infinite-dimensional space in the theory of integral equations.

In the axiomatic treatment of the linear vector space, and in particular of the Hilbert space, nothing is required but that the vectors can be added together and multiplied by numbers, the main axiom being the vector product one.

In that case the vector space may be a wide variety of element sets. So, the set of functions specified on a segment also form a linear vector space, if the scalar product is represented as the integral of their product

$$(x, y) = \int_a^b x(t)\, y(t)\, dt. \qquad (**)$$

To be sure, this space does not include all the conceivable functions, but only those for which there exists an integral of the square of the function, the square of the vector length

$$(x,\ x) = \| x \|^2 = \int_a^b x^2 (t)\, dt.$$

The space of all such functions is also called a Hilbert space and denoted by $L_2 (a,\ b)$.

Note now that if two vectors x_1 and x_2 in a vector space are not parallel, then the set of all their linear combinations $a_1x_1 + a_2x_2$, where a_1 and a_2 are arbitrary numbers, ills the plane. Accordingly, linear combinations of n vectors x_i of the form $a_1x_1 + a_2x_2 + \ . \ + a_nx_n$, where a_i are any real numbers, generally fill all the n-dimensional space. It is said to be spanned to vectors $x_1,\ x_2,\ .\ .,\ x_n$.

If the dimensionality of the initial space is more than n, then the n-dimensional space obtained is called a subspace of the initial space. In a Hilbert space, an infinite-dimensional one, any space spanned to its n vectors will be a subspace. However, in a Hilbert space there are infinite-dimensional subspaces as well, e.g. the one spanned to all the unit vectors, besides the first three, or spanned to all the position vectors with odd numbers.

We will now consider one simple problem of elementary geometry. Let in a conventional three-dimensional space R there be a plane Q passing through the origin of coordinates and a vector y not belonging to that plane. How can we find in Q a vector lying the closest to y? To be

sure, you know the answer: if from the end of y we drop the perpendicular to plane Q, the resultant vector $\hat{y}$, the projection of y on Q, will be the closest to y among all the vectors in the plane. In other words, the best approximation of y with the help of the vectors in Q will be $\hat{y}$ (Fig. 15).

A similar problem of the best approximation also occurs in the theory of multidimensional

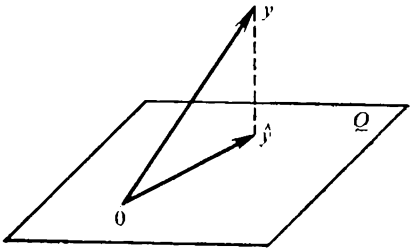


Fig. 15

spaces: if y is a vector in a space H (of any dimensionality) and Q is its subspace that does not contain y, then the best approximation of y by vectors from Q will be $\hat{y}$—the projection of y on Q.

We can even obtain the error of the best approximation: it is obviously the length $\| y - \hat{y} \|$ of the perpendicular dropped from the end of y to subspace Q.

If subspace Q is spanned to the vectors x_1, x_2, ., x_m, then to drop the perpendicular from the end of y to Q is to find a vector z that is orthogonal to each of x_1, x_2, ., x_m. Such a problem easily comes down to solving a system of linear algebraic equations.

In our geometrical interpretation everything appears quite simple. But remember now about the Hilbert space of functions. Here the vectors are functions specified on a segment, a subspace spanned to n vectors—all the possible combinations of these functions, and so the problem reduces to finding the best approximation to a certain function using the above linear combinations. Analytically, the best approximation problem for a given function by the linear combinations of other functions does not appear to be that simple, and the geometrical treatment indicates one of the possible ways of solving it, providing quite a lucid picture of all the operations necessary.

The above problem of best approximation is other than in the Weierstrass theorem, the formulation of the problem of approximating a function by a linear combination of some simpler functions is simpler in the sense that in various formulations different treatments of the distance between the functions are used. In the Weierstrass theorem the distance between functions is taken to be the largest distance between their curves along the vertical axis, or rather its magnitude. But here the distance between vector-functions is the norm of their difference

$$\| x - y \| = \sqrt{\int_a^b [x(t) - y(t)]^2 \, dt},$$

i.e. the square root of the surface area between the horizontal straight line and the curve of the squared difference of the functions.

When the problem of representing a function in the form of a sum of some "standard functions" could be viewed from some more general aspect, it became clear that there are many systems of functions from which, just as from building blocks, we can construct any continuous function.

The initial functions to be used for approximating the given function can conveniently be a sequence of mutually orthogonal functions—the orthogonal basis of the Hilbert space of functions under consideration.

This implies that any function from the space can be represented by a linear combination of the functions of the basis. The basis functions are thus the building blocks of which all the variety of the functional space in question is composed.

Sunrise, Sunset.

Day and night alternate, seasons alternate, the heart contracts, say, seventy times a minute...

These all are alternating, periodic processes. Their periods are a day, a year or 1/70 of a minute. Such periodic processes are to be encountered everywhere. It has been found, for example, more than a hundred of physiological systems functioning with diurnal periodicity.

One of the oldest technical devices based on the simple periodic motion is the pendulum. Shift it aside and let go of it and the pendulum will swing to and fro. We will ignore now the friction that dampens its motion and watch the time variation of α, the angle of deflection of the

pendulum from the equilibrium, a vertical line. You will know from your school physics course that α is a periodic function of time, and its period is the period of pendulum oscillations, i.e. the time taken by the pendulum to come back to its extreme left position, say. We can describe the motion of the pendulum by the formula

$$\alpha(t) = A \cos\left(\frac{2\pi}{T} t + \varphi\right),$$

where A is the amplitude of oscillations, t is the time, T is the period of oscillations, φ is the initial phase. They normally use circular frequency

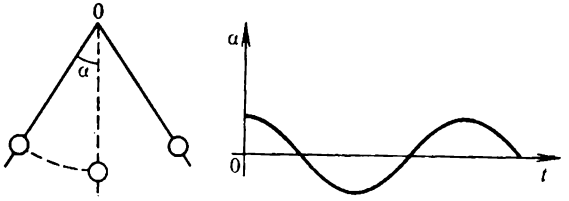


Fig. 16

$\omega = 2\pi/T$ and then the curve of Fig. 16 will be given by $\alpha(t) = A \cos(\omega t + \varphi)$.

There is an infinite variety of periodic processes. Let us take some examples from various fields.

When in a lake there live plant and crustacean-eating fish and fish-eating fish (e.g. pikes), and the former are plentiful, the fish-eaters have much food and multiply prolifically. As a result, the plant-eaters decline drastically, while the population of fish-eaters explodes, and so gradually the latter will suffer from shortage of food. Now the population of fish-eating fish drops and that

of plant-eating fish shots up. Again the fish-eaters have food in large supply, and so the situation recurs.

Similar periodical processes also occur in economy. In free enterprise economies there are cyclic variations of prices of agricultural products. So premium prices of pork stimulate farmers to rear more pigs. And so, according to a German economist, in the 1920s the supply of pork would increase dramatically in the market

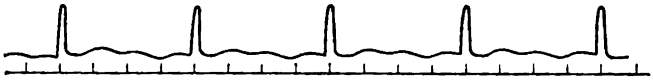


Fig. 17

in about 1.5 years, so that the prices were brought down. The process reversed: the farmers cut pork production up to the moment when, now again due to the shortage of pigs, the prices soared again. If no other factors intervened, the pork price underwent fluctuations that were about sinusoidal in character with a period of about three years.

A number of periodic processes are closely related to systoles. Many people know of electrocardiograms, a record of biocurrents picked up at a region close to the heart. Figure 17 shows an electrocardiogram clearly displaying a periodicity of peaks, current pulses.

It would not pay to use polynomials to approximate periodic functions since too high powers would be required, and so here the use is made of simple harmonic motions at frequencies that are integral multiples of the fundamental frequency. So if the period is T and the circular

frequency is $\omega = 2\pi/T$, then the building blocks will be the sinuosidal functions of multiple frequencies: 1, $\sin \omega t$, $\cos \omega t$, $\sin 2\omega t$, $\cos 2\omega t$, ., $\sin n\omega t$, $\cos n\omega t$. We thus obtain trigonometric polynomials of the form

$$s(t) = 5 - 2 \sin t + 0.3 \cos 2t - 0.1 \cos 4t.$$

Notice that the coefficients here are arbitrary, and the frequency is $\omega = 1$.

In the general form, the trigonometric polynomial will be

$$s(t) = a_0 + a_1 \cos \omega t + b_1 \sin \omega t + \\ + a_n \cos n\omega t + b_n \sin n\omega t,$$

where a_0, a_1, b_1, a_n, b_n are numerical coefficients.

It appears that any continuous periodic function can within any accuracy be approximated by a trigonometric polynomial.

This theorem, as fundamental as the previous one, is also due to Karl Weierstrass. Although as early as the first quarter of the 19th century Joseph Fourier (1768-1830) in his studies of thermal conductivity made effective and efficient use of representation of functions in the form of sums of sinusoidal oscillations of multiple frequencies. Therefore, series representing functions by sums of simple harmonic motions are referred to as Fourier series.

If now we apply the geometric approach discussed in the previous section, we will find that we here deal with the same problem of representing a function using linear combinations of basis functions, and the latter in this case are trigonometric functions of multiple frequencies. We

will only have to make sure that these are orthogonal. To this end, we will take some simple integrals.

Fourier coefficients are scalar products of the initial function by basis trigonometric functions. They are expressed by integrals in a simple way. These standard formulas are to be found in any calculus manual.

Note, in passing, that the power functions x^m discussed in section "Building Blocks" are not pairwise orthogonal. But among the polynomials we can also select some pairwise orthogonal ones. Orthogonal polynomials were first introduced in 1785 by the outstanding French mathematician Adrien-Marie Legendre (1752-1833). I will only give the first five of the Legendre polynomials:

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1),$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x),$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3).$$

It is easily seen that they are orthogonal within the segment $[-1, +1]$, i.e. at $n \neq m$ we have

$$\int_{-1}^{+1} P_n(x)\, P_m(x)\, dx = 0.$$

If we were to give an absolutely precise representation of a continuous function by an algebraic sum of basis functions, then the series would generally contain an infinite number of terms.

Such a representation would, of course, rejoice the heart of a pure mathematician, and indeed it is beautiful. But the practician will have to contrive to represent his complicated function within the accuracy required by a sum of a few simple functions.

If we have to represent within (a, b) a continuous function to within δ, then the problem can be solved by taking, for example, in our approximation the first n terms of its Fourier expansion. But such a representation may also contain a fairly large number of functions, it grows with decreasing δ. Therefore, an adequate selection of basis functions to provide in the problem at hand the satisfactory accuracy of approximation by a small number of basis functions is a point of vital importance here.

Now I think you expect some recipes or at least recommendations as to how to select a basis that provides an adequate accuracy by using a small number of simple functions. Unfortunately, as is often in applied mathematics, no general recommendations can be given—the success in selecting the basis functions is conditioned by the nature of the problem in question, the information about the object studied and experimental evidence available. If, say, the function of interest to us is a response of a linear system with constant parameters (a passive two-port in the language of radio engineers), then the basis must, of course, be sought among simple harmonic motions and exponents, if we deal with the response of a system with varying parameters (e.g. the response of an oscillatory circuit with varying capacitance), then the basis will be special func-

tions that essentially depend on the law governing the variation of the parameters. If you are not acquainted with them, they will appear rather difficult to you.

But back to the cloud of data of Fig. 13 in the section "Regression" We do not know the regression equation $y = \varphi(x)$ and we cannot derive it from the experimental evidence, however extensive. Recall that at first we should suggest a hypothesis as to the form of dependence or, put another way, to think of a mathematical model, and only then test it drawing on experimental data.

Now we chose the model gazing at the cloud of points and resurrecting a swarm of associations. To be sure, we begin with a simple model, e.g. a linear model (i.e. a straight line), a sinusoid, a parabola, and so on.

I have read somewhere the following: "One of the principal tasks of a theoretical investigation in any field is finding a point of view such that the object in question appears in the simplest way". This is, of course, true, if only the difference between the simple and the complex is clear. But simplicity is a fairly arbitrary entity, being substantially conditioned by habits, experience, and knowledge. I was told that a seventy-year old was treated to a pie with seventy candles at his birthday party. He offered to his three-year old grand-daughter to blow the candles out and the kid asked, "But where is the switch?"

A secondary school student may be dismayed by sines and cosines, especially if these have been introduced in a formal way, as ratios of sides to the hypotenuse, without explaining their re-

lation to oscillations. At the same time, the electrician is used to sinusoidal 50 Hz oscillations, since this is a feature of industrial alternating current used in your household. The TV repairman deals with pulsed periodic processes (see, for examples the plots in Fig. 18).

Therefore, if in a problem even having nothing to do with television we select for the basis some

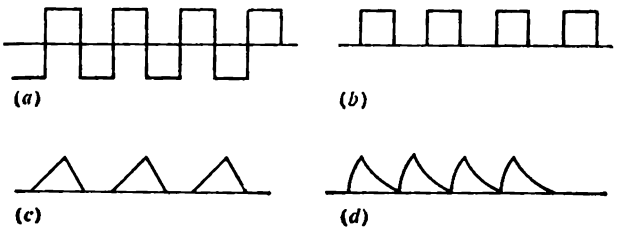


Fig. 18.

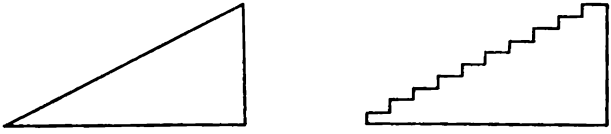
pulses with different repetition frequencies and different amplitudes, a TV man will not be surprised by this representation, since it is natural for the class of processes occurring in television.

But a basis of rectangular pulses should not be used to represent triangular pulses in Fig. 18*c*, since such a representation, even with low accuracy, would require a model of a large number of basis functions. As is shown in Fig. 19, the triangular pulse (Fig. 19*a*) is approximated by the sum of ten pulses of the same duration but different amplitude (Fig. 19*b*) and the resultant ladder is still a poor approximation.

Selection of good basis thus requires both experience and clear understanding of the physical

nature of the problem at hand. These would help throughout the procedure, though.

But the laws of nature or actual dependences between variables in a reasonable model in physics, chemistry, biology, economics, sociology are generally not very complex. If factual evidence



(*a*) (*b*)

Fig. 19

leads to a model that is only satisfactorily described using a polynomial of the hundredth order, then common sense dictates that it is necessary either to revise the formulation of the problem or select other system of basis functions. This means that another model is required, which may be combined from other typical functions corresponding both to the object and to the problem.

My belief in the simplicity and clarity of the laws of nature and actual dependences even in complicated systems is based on experience. And if you ever constructed models of real-life processes or objects, you will agree with me; and if you are to do this in future, then a measure of scepticism in relation to ornate, obscure and unwieldy reasoning would be a good guidance, and I hope such scepticism will do you much good.

In short, before plunging into calculations or an experiment give some thought to the selection of a mathematical model—this may exert a crucial influence on the final result.

The Nearest

Now you have decided on a model and can turn to the coefficients.

For example, for the cloud of experimental points in Fig. 13, which gives some relationship between x and y, it seems reasonable to try the model in the form of the third-order polynomial

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3,$$

where α_i are numbers to be found, or rather estimated from the experimental evidence available.

Of course, we can plot a whole family of curves for similar polynomials, so that they will adequately correspond to the cloud of Fig. 13. Such plots are shown in Fig. 20. In other words, there are many ways of choosing the coefficients so that the polynomial would give a good approximation of that unknown dependence, which, we hope, does exist. The spreads graphically represented in the figure can naturally be regarded as a result of chance, they may be due to measurement errors, "errors" of nature or other causes.

But we can hardly be happy with such a variety, because if there is in fact some functional dependence $y = f(x)$ between x and y, then among all the mathematical models of a selected type (in the example under discussion, among all the polynomials of the third order) there must be one

that comes the closest to $f(x)$. How are we to find this best model among all the models of the type chosen?

The reader who is versed in such things or who at least has paid some attention to the section "A Glimpse of Criteria" will recall that it is necessary at first to select a measure of closeness

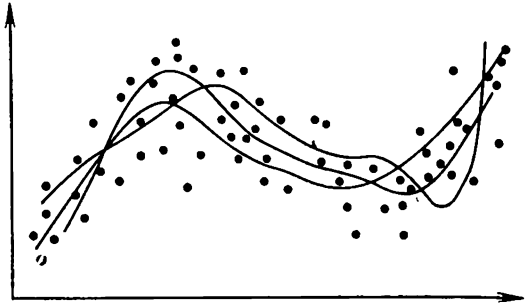


Fig. 20

of sticking of one function to the other, a criterion of closeness of the two functions. The discussion of the various measures possible is beyond the scope of the book, instead we will consider some fruitful idea of choosing the most suitable model.

Notice that the function $f(x)$, which we want to reproduce, if only approximately, is unknown. All the information about it is contained in the cloud of experimental data and those considerations on which the model chosen is based.

If we draw the curve of one of the possible realizations of the model of the type chosen, i.e. at some definite numerical values of the coefficients, then the curve will pass in some way or

other among the points of the cloud. We will now connect each experimental point with the curve by a segment parallel to the coordinate axis, as in Fig. 21. The collection of these small segments shows to what extent the curve drawn corresponds to the experimental points.

But the segments are many and they differ in length, and so we will have to think of a way

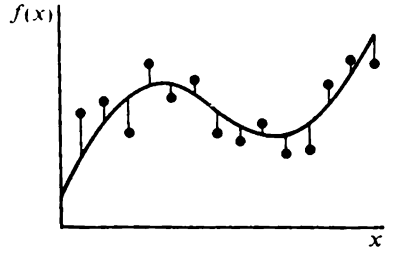


Fig. 21

of using the ensemble of these segments to turn it into one number, a criterion.

For the ordinate axis the upward direction is positive, so that the length of a segment from the point to the curve will be positive when the point lies above the curve, and negative when the point lies under the curve. The algebraic sum of the segments, therefore, does not characterize the quantity of interest to us, and so we want something better.

We may take the sum of lengths of the segments without the signs, i.e. the sum of their magnitudes, but a much more convenient measure is the sum of squares of the lengths. Accordingly we must choose a model, or rather those

coefficients α_0, α_1, α_2, α_3, . . at which the sum of squares of lengths will be minimal.

You may well ask why we should use the sum of squares, rather than the sum of fourth powers or some other function of lengths.

It was Legendre who suggested in 1805 in his article "New Methods for the Determination of Comet Orbits" to use the sum of squares (the least squares method). He wrote: "After all the conditions of the problem are satisfied, it is necessary to determine the coefficients so that the errors would be the smallest of the possible. To this end, we have developed a generally simple method consisting in finding the minimum of the sum of squares of the errors."

You see thus that Legendre did not choose to explain why he had selected the sum of squares. However, behind this selection there are rather profound ideas. The method of selecting the coefficients for a model based on minimization of the sum of squares of deviations is called the *method of least squares.*

But soon the great Gauss in a number of works gave a probabilistic justification of the method of least squares. In the earliest works the method was closely linked to the normal distribution of measurement errors and Gauss justified the method using the concept of maximum likelihood. But the most critical properties of estimates of the coefficients turned out to be independent of distribution.

We will consider the Gauss's approach with reference to the same cubic regression equation. We will denote the estimated values of α_0, α_1, α_2, and α_3 by a_0, a_1, a_2 and a_3, respectively. Observa-

tions yielded n pairs of values $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, the corresponding points in the plane (x, y) forming the cloud. We will assume that $x_1, x_2, \ldots, x_n$ are determined without error: in physical, economic or engineering problems the variable x is often defined by the investigator, and so it may be time, established temperature or whatever. Thus, the random deviations from the desired precise dependence

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$$

for each point (x_i, y_i) are vertical segments in Fig. 21, given by

$$\delta_i = y_i - (\alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 x_i^3).$$

We will now treat the finding of a_0, a_1, a_2, and a_3 as a hazardous game in which you cannot win but can only lose. We will take the measure of loss to be $\sum \delta_i^2$, so that the loss will be the larger the larger the uncertainties δ_i, i.e. the lengths of the segments.

Let us now formulate the main requirements. First, the estimates must not contain systematic errors, i.e. the mathematical expectation of a_k must be equal to α_k. This also means that expectations of each of δ_i must be zero. Second, the expectation of the total squares of losses, i.e. the variance of the total error, must be the smallest among all the other estimates.

It turned out that the estimate meeting these two requirements yields exactly the coefficients derived by the method of least squares.

Contributors to the method of least squares are Laplace, Chebyshev, and Markov. The latter gave a consistent generalization of Gauss's re-

sults, so that now the results are generally known as the Gauss-Markov theorem.

Computation using the method of least squares comes down to solving a system of linear algebraic equations. There are no principal difficulties here but there are some notable achievements in computational procedures, which however lie beyond the scope of the book.

The Art of Hope

A boxing champion running into a stray dog stops: although the boxer is far stronger than the dog, he does not want to be attacked by it. Of course, he is optimistic and neither touches the dog nor seeks refuge in a doorway, but the boxer cannot be sure of the dog's actions and waits gathering information to predict its behaviour. A tamer, too, is optimistic when he pokes his head into a lion's mouth, whereas the public still do not exclude the possibility of a disaster, otherwise there would be nothing to stun them.

The Russian psychologist V. Levi wrote about the art of hope in his book *Hunt for Thought*: "We are the product of the flexible, mobile, multidimensional world of living nature swarming with versions, overflowing with possibilities and contingencies. There is little in this world on which you can rely completely, and it is exactly for this reason that living things had to learn the art of hope. It was a hard school. Those who had hoped poorly died the first.

"To hope well means to hope incessantly. And not rigidly and dully, but flexibly. Not blindly,

but vigilantly and correctly. To hope well means to choose well what you can rely on. This means to be able to change you choice in time. This means to be able to weigh chances, assess probabilities. In short, to be able to foresee and rely on the foresight.

"The art of hope is the art of achieving a goal. But goals may differ widely.

"The art of hope is necessary to swallow diving to intercept a midge, and for a goal-keeper waiting for the ball in the right angle of the goal, and for a gunman aiming at a flying target."

When a girl tells fortunes by a camomile—"he loves me, he loves me not"—or listens to the muttering of a Gipsy, and bases her hopes on this, she does not master the art of hope.

All superstitions are based on the same principle: hope proceeds from random coincidences, facts not related by some cause-effect ties. Consider a funny example from modern sporting life.

A day before the final match of the 1976 USSR Football Cup a football expert wrote in the newspaper *Sovietsky Sport* (Soviet Sports) of 3 September:

"I always like to search for some indications that could prompt something as to the outcome of the forthcoming finals. So I put forward the supposition that the winning team will be the one that will be the first to score a goal. At any rate, in recent years the following trend has been observed: in odd years the winner was the team that was the first to concede a goal and in the even years the one that was the first to score."

Next day at the final match the first scorer won the cup. The year being even, the author's hy-

pothesis was verified. But the author, I think, only put forward his "hypothesis" as a joke, because the situation here is characterized by the total lack of any cause-effect relations between the outcome and the oddness of the year.

And still coins or dice continue to be used to tell fortunes and the results are claimed to bear some import. But the occurrence of head or tail now exerts no influence whatsoever on the chances of having, say, a head in the next toss for a model with equiprobable and independent outcomes. Should the head appear repeatedly several times, a gambler may become suspicious of the very model and of the assumptions that the model is equiprobable and independent, although such a situation is possible in principle.

I have repeatedly performed this experiment with my students: if, say, in coin tossing head (or tail) appeared seven, or ten times in succession, then nearly all the audience would vote for a revision of the equiprobability assumption and often suspected some cheating on my side, and with good grounds.

In actuality, a sequence of independent trials provides no information as to the future of equiprobable outcomes and only meagre information with unequal probabilities, the information being the richer the larger the difference between the probabilities of outcomes. The above reasoning may be viewed as appeal to use common sense in playing head or tail, dice, or roulette.

But the book is no guidance as to how to behave in a gambling-house. In real life events are greatly influenced by previous events, and it is on this dependence that the art of hope relies.

Newton's laws enabled planetary motions to be described precisely. Real gravity forces deviate but slightly from the famous law. The travel of a shell fired from a gun can also be predicted with sufficient accuracy, but still much lower than the motion of the planets. Here a number of factors are involved that affect the trajectory of the shell's motion as compared with the calculated one: fluctuations of weights of the shell and explosive, distortions of the shape of the shell and the gun, atmospheric and gravitational inhomogeneities, and so on. As a result, shells fired at a long distance hit not very often, they explode in the vicinity of the target. But it should be stressed that the trajectory can be predicted fairly accurately and for skilled gunners hitting a stationary target is a long solved problem. With a moving target the situation is different.

You will have tried to shoot a running game or hit with a ball a scuttling boy in some children's game and you know that it is necessary to aim with a lead. But you do not know for sure how the target is going to move from now, and so you just hope that the target will get to where you aim.

Many misses in our life, in the direct and figurative meanings, show unfortunately that the art of hitting a moving, changing target—the art of hope—does not come easy, and even great experience does not save from failures.

In 1948 Norbert Wiener published his book *Cybernetics: or, Control and Communication in the Animal and the Machine* that revolutionized our ideas of control. One of the problems that Wiener laid at the foundation of general concepts of cy-

bernetics was gunfire control. Wiener's reasoning was used as a basis of very many later works on control in situations where some prediction of the object under study is necessary. These ideas were further developed in his later book *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (1949).

The reasoning was as follows: if a plane moves steadily and along a straight line, then from the radar indications of its location and velocity, the point could be predicted where it would be in time τ required by an antiaircraft shell to cover the distance from the gun to the aircraft, and so to aim the gun to that point. But the plane does not move steadily and along a straight line even during a short time τ and its trajectory is random and not amenable to unique prediction.

But examination of recorded trajectories makes it possible to construct a mathematical model of the trajectory, and for the model Wiener took the stationary random process. Roughly speaking, stationarity implies that the stochastic behaviour of the process is homogeneous in time.

Recall building blocks from which desired functions can be constructed. In terms of them, it is convenient to view a sum of simple harmonic motions of various frequencies (not necessarily multiples of some fundamental frequency), whose amplitudes are independent random variables. Any stationary random process can, to within any accuracy, be approximated by a similar trigonometric sum.

Let us now return to the problem. The gunfire controller must determine the direction of the point where the shell will intercept the target

and wait for the command. The difficulty here is that the trajectory of the target is random, since each coordinate (e.g. the distance and azimuthal angles of the target) is one of the many realizations of a stationary random process. Consider any one of the coordinates. Let the signal

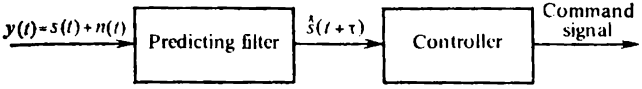


Fig. 22

coming from the radar to the controller be $y(t)$, it representing a combination of the signal corresponding to the observed coordinate of the target trajectory $s(t)$ (called the legitimate signal) and the noise $n(t)$. Noise is always present—it is the receiver noise, atmospherics, and so on. To a reasonable accuracy, such noise can also be considered realizations of some stationary random process, which is, of course, different from the legitimate signal and statistically independent from it. Thus, to the controller (Fig. 22) comes the combination

$$y(t) = s(t) + n(t).$$

The task of the first block—the predicting filter—is to produce a signal $\hat{s}(t + \tau)$ that is as close as possible to the real value of the signal $s(t + \tau)$ at the time $t + \tau$ of meeting.

In other words, we must select the signal $\hat{s}(t + \tau)$ providing the best approximation to the real coordinates of the target in the time τ the shell takes to reach the target.

It is worth stressing the specific features of the approach just described. If the signal $s(t)$ (and hence the target trajectory) were known for sure, the problem would present no difficulties in principle, as it would reduce to some algebra. If the noise $n(t)$ were known exactly, then, since the observed signal $y(t)$ is known and $s(t) = y(t) - n(t)$, then $s(t)$ would also be determined exactly, and the problem would again be trivial. In the other extreme situation, when a priori there is no information about $s(t)$ and $n(t)$, there also is no hope of having a good estimate proceeding from $s(t) + n(t)$ alone. The latter situation is like die tossing, where the future is in no way conditioned by the past. Engineers, however, know a thing or two about the structure of noise and about possible trajectories of airplanes or other targets, and it is this knowledge that enables them to work out a model in question.

Now we must, of course, refine the criterion of quality of prediction or the criterion of closeness between the actual value of the signal $s(t + \tau)$ and the estimate $\hat{s}(t + \tau)$ predicted by the filter. As we have repeatedly mentioned earlier in the book, the criteria to be suggested are legion. But if we are to stay within the logic of least squares, we should assume as the closeness criterion for $s(t + \tau)$ and $\hat{s}(t + \tau)$ the expectation of the square of their difference

$$\rho = \mathsf{M}[s(t+\tau) - \hat{s}(t+\tau)]^2.$$

The selection of the best prediction will then reduce to the selection of the filter to minimize

ρ. It is worth noting, of course, that it is not always that such filter will give the best prediction, since it only gives the values of $\hat{s}(t+\tau)$ that yield the best prediction of average only. But here we are in the realm of randomness, and so we could hardly come up with something that would always provide the optimal prediction.

The problem of the best prediction of the behaviour of a stationary random process lends itself to a graphic geometric interpretation.

The general concept of the linear vector space enables us to view the ensemble of random variables (to be denoted by Greek letters) with zero expectation and limited variance as a Hilbert space, such that the scalar product of the random variables—the vectors in our space—is the expectation of their product

$$(\xi, \eta) = \mathbf{M}\xi\eta,$$

and the square of the vector length is its variance. The distance between the two vectors ξ and η will then be given by

$$\| \xi - \eta \| = \sqrt{\mathbf{M}(\xi - \eta)^2}.$$

Let $\xi(t)$ be a random process. At each time t for each random variable $\xi(t)$ we can find a vector in the Hilbert space H of random variables. The random variable $\xi(t)$ varies with time t, i.e. its vector goes over to another position. As the time t passes from a to b, the end of the vector in H describes a curve, and on the set of its vectors a subspace, say $H(a, b)$, can be spanned.

Let now t be the observation time and we are interested in the value of a random process at

some future time $t + \tau$. Generally speaking, it is impossible to predict $\xi(t + \tau)$ uniquely, and so we should be happy with an optimal estimate of these values. Geometric interpretation brings out the solution at once. Let us construct the subspace $H(a, t)$ so that it answers to the past (up to the time t) values of the random process. If $\tau > 0$, then the vector $\xi(t + \tau)$ is the "future" of the process $\xi(t)$ at time $t + \tau$ and it, generally speaking, does not enter the "past" subspace $H(a, t)$. Otherwise, exact prediction would be possible. The best approximation for the future value $\xi(t + \tau)$ in the interval (a, t) will be the projection of $\xi(t + \tau)$ on the past subspace $H(a, t)$, and the length of the perpendicular from the end of $\xi(t + \tau)$ on $H(a, t)$ will be equal to the error of prediction.

If $\xi(t)$ is a stationary process that behaves in a way common for practical applications, then these geometric arguments enable relevant expressions for the prediction to be written and even appliances, such as predicting filters, to be constructed to realize the best prediction.

The mathematics of determining the behaviour of the best linear predicting filter, which was suggested by Wiener, is rather complex, drawing on available methods of functional analysis, integral equations, and functions of complex variable, which lie beyond the scope of the book. It is only worth noting here that the prediction methods for the theory of stationary random processes were developed in 1939-1941 by the prominent Soviet mathematician Kolmogorov, but it was Wiener who not only put forward his theory independently, but also applied it to an impor-

tant technical problem. Now a vast body of literature on random processes is available. It covers weather forecasts, autopilot control, blind landing of aircraft, marketing and logistics, short and long-term planning in economics, and what not.

The methods originated by Wiener and Kolmogorov are tools for studying a wide variety of problems in radiophysics and radio engineering, atmospheric physics, control theory and so on. As is often in science, a mathematical tool that has come from physics, engineering, biology or other branches of science at a later date finds uses in other fields.

A Soviet mathematician, A. Khinchin, worked out the mathematical tool to treat a class of stationary random processes in 1935. In the 1940s his technique was effectively employed in radio engineering to solve the problem of filtration, i.e. separation of the transmitted signal from noise. The process is essentially like this. To the input of a receiving device comes a deterministic useful signal $s(t)$ and noise $n(t)$, which is a stationary random process. The receiver must single out the signal while suppressing the noise as far as possible. The task at first was to calculate the signal-to-noise ratio at the input. Later on, the problem became more complicated: it was necessary to select the waveform and receiver response to maximize the signal-to-noise ratio at the output. At later stages, the problem was formulated in a more general way: to select the receiver response so that to optimize, according to some criterion, the relation between signal and noise characteristics at the receiver

output. In the process, not only noise $n(t)$ but signal $s(t)$ as well are treated as stationary random processes. The situation where $s(t)$ is a deterministic signal may be viewed as a special case of a stationary process. Now we can make use of the Wiener-Kolmogorov theorem, assuming that the prediction time is $\tau = 0$ and that the problem comes down to designing a receiver (mathematically, in selecting an operator) such that the output signal will, in terms of the method of least squares, differ from the useful signal at the input as little as possible, i.e. when the expectation of the square of the signal deviation at the output from the useful signal at the input will be minimal. The problem is solved using the same methods as the above-mentioned prediction problem.

Struggling for a Record

The previous section dealt with the prediction of future positions of a moving object. We will now discuss other prediction problems.

An ambitious sprinter puts in for top results in 100-metre run. His personal record is 10.4 seconds, and he is eager to cut at least two tenths of a second. But these last tenths of a second come especially difficult, and hence a severe routine is necessary. Clearly, the result is a function of time spent on training and the diet, both can be varied within reasonable limits. So a happy mean is necessary. For one thing, the runner should take care of his weight, and hence the number of

calories consumed, on the other hand, he should have enough energy for his training, i.e. enough calories. So the result is a function of the daily consumption of calories, both overnutrition and undernutrition being bad for his performance. It looks like the plot of the variation of 100-metre run time with the number of calories is parabola-like, i.e. has one minimum.

If the sportsman puts in insufficient time into his training, he will not be in good shape. On the other hand, if he puts in 15 hours a day, again no good results can be expected—fatigue will result in flabbiness, indifference—no records for you. Therefore, the result in terms of seconds will again vary with the duration of daily training sessions as a parabola.

It should be stressed that our reasoning suggests that there exists the best training system and it only remains to find it. To be sure, training requirements are highly specific and there does not exist a solution common for all the sportsmen. Therefore, the search for the best procedure must only be based on observations specific for a given runner.

We will now make use of the ideas of regression analysis. The result τ is postulated to be a function of two variables or factors, the duration of daily training sessions, t, and the number of calories daily consumed by the sprinter, v:

$$\tau = f(t, v).$$

The form of this function is unknown but we may well assume that it is a paraboloid given by the equation

$$\tau = \beta_0 + \beta_1 t + \beta_2 v + \beta_{11} t^2 + \beta_{12} tv + \beta_{22} v^2. \quad (\beta)$$

This is a regression equation, which only differs from those considered above in that it is in two independent variables t and v, not one.

If we knew the coefficients β_0, β_1, ., β_{22}, then we could easily design the training system for best results, i.e. find the pertinent values of t and v. In other words, we could determine the coordinates t_0, v_0 of the lowest point τ_0 of the paraboloid. The coefficients being unknown, the sprinter and his coach can only make observations to be used to work out approximate values for the coefficients and hence approximate values for t_0 and v_0.

Each experiment takes a long time, actually several weeks, for the sprinter's organism to adapt to the new procedure and for the results to be obtained for this procedure. Therefore, it is impossible to try very many forms of the procedure but it is desirable to arrive at the optimal one as quickly as possible. What is more, for a chosen procedure we should admit some unpredictable things: the sprinter may get not enough sleep, he may be under some stress, the weather may be uncooperative, and so on. In short, there are many impacts not covered by the model selected that may predetermine the spread of results for the same training procedure. We will take this spread to be random, and so will make use of statistical techniques in estimating the coefficients. The shrewd reader may now have surmised what I am driving at: the coefficients are best estimated using the method of least squares. In the previous sections the method of least squares was used for functions in one variable, but several variables change essentially nothing both in con-

cepts and in the formalism: we again must solve a system of algebraic equations for the unknown coefficients β_0, β_1, ., β_{22}. The only knowns here are observation results: for each given training procedure (t_i, v_i) the sportsman will make several 100-metre runs and show the results $\tau_i^{(1)}$, $\tau_i^{(2)}$, $\tau_i^{(m)}$, which may generally be different.

Geometrically, the picture is like this: on a plane points are marked—the procedures tested—and over each of them we locate a point whose vertical coordinate is the result shown.

Now we will have to construct such a surface of paraboloid (mathematical model) that would best correspond to the points (t, v, τ) obtained, and then on this surface we find the minimum τ_0 and its coordinates (t_0, v_0), which will be taken to be the best training procedure.

Again the shrewd reader will have noticed that we in fact have here two problems at once: the identification, i.e. the construction of a model from observation results, and the optimization, i.e. the selection of the best procedure. Let us analyse these problems one after the other and take note of what we have done to solve each of them and what conclusions are to be drawn from the above discussion. At first the identification.

We know the sprinter's performance for some of the observed procedures. The guy and his coach are, of course, interested in that eternal question, "What would be if ...?"—in this case if they would choose other training schedules and diets, for example at 4,800 calories and 5 hours of training a day. If a mathematical model is satisfac-

torily represented by the experimental evidence available, when the model is said to be adequate, then it provides the answers. To this end, into the expression

$$\tau = b_0 + b_1 t + b_2 v + b_{11} t^2 + b_{12} tv + b_{22} v^2 \quad \text{(b)}$$

we will have to substitute the values $\bar{t}$ and $\bar{v}$ corresponding to appropriate procedure and, after some algebra, to arrive at $\tau = \tau(\bar{t}, \bar{v})$. Here $\bar{t}$ and $\bar{v}$ are fixed values: $\bar{t} = 5$ hours, $\bar{v} = 4{,}800$ calories.

Note that in the two last formulas (β) and (b), which are similar in form, the coefficients are denoted by different, but again similar, letters. This is not without purpose: in the first expression the coefficients are some numbers, and in (b) coefficients are found by the method of least squares, and they are thus estimates of appropriate coefficients in (β). Such a system of notation is quite common in regression analysis.

But if instead of exact values of coefficients we substitute their estimates, then instead of the value of τ we will only have its estimate, i.e. its approximation. The estimate will be the more accurate the less is the number of observation points on which it is based.

Consequently, from experimental data it is possible to construct the regression equation (b) and use it to predict the values of the parameter of interest (in this case, the time τ taken by the sprinter to cover the 100 metres) at the points (procedures) lying within the region under study, the accuracy being the higher the larger the number of observation points used in constructing the regression equation.

Does the accuracy of prediction here depend on the arrangement of the points where observations have already been made? The answer is by no means self-evident and calls for some speculation.

It is to be noted that in discussing the sportsman's routine we have only selected two factors—the duration of training and the number of calories. But, of course, the condition of the sprinter and his potentialities are dependent on other factors as well. Say, if his training session lasts 6 hours, these may be put in from 9 a.m. to 3 p.m. without a break, or these can be split into two sessions three hours each, or else into three two-hour sessions, obviously, with different results.

Besides, trainings themselves may be quite different: each sportsman should receive comprehensive training, so that a runner's training should include a good deal of other track and fields, gymnastics and weight-lifting.

As far as the diet is concerned, it is not only calories that matter, the diet should be varied, including proteins, fats and hydrocarbons, vitamins and trace elements, and the percentages of these can be different. Lastly, apart from training and diet, many other factors determine the condition of a sportsman: sleeping hours, age, and so on, which, although not expressible in numbers, are very important. For example, the morale or the general cultural level.

Thus, the daily routine and the general condition of a sportsman are described by a fairly large number of factors, thus complicating the problem of selecting the optimal training procedure.

Let us leave our sportsman for a while and turn to the problem of selecting an optimal working condition for a functioning plant at a factory.

The observed factors here will be the rated parameters of the process, e.g. temperature, pressure, flow rates, concentrations of substances involved in physical and chemical reactions. Under normal operation, their values are not set arbitrarily.

An example considered in some detail above was the electric desalination process (EDP). But we have only discussed the input-output relation, or rather the amounts of salts at the input and output of an EDP. In actuality, however, the efficiency of desalination process is dependent on many factors, say, on the temperature of raw material, the amounts of flushing water and demulsifier, the electric field strength, and the time of residence of the emulsion in the electric field.

With more than two factors involved, the graphic representations become impossible, although, as is said above, both the associations and terminology remain: we now speak about a multidimensional factor space and a surface in that space that reflects the dependence of the parameter of interest (e.g. the amount of salts at the EDP output or the 100-metre run results) on all the factors involved. This surface in the theory of experiment is referred to as the *response surface*. What is meant here is the response of the object in question to the input in the form of some set of factors.

In the case of identification the task is thus to construct the equation of the response surface

(it is called the response function) and to test the suitability, or adequacy, of the equation. Should there be no adequacy, everything must be iterated: think of some other mathematical model, test its adequacy, etc., until adequacy is achieved. It is only then that the equation derived can be used to predict the values of response at points within the region studied, which are, of course, different from those used to estimate the regression coefficients.

Vices of Passiveness

Observations of cometary motions, technological processes under normal operation, achievements of a sportsman with arbitrarily selected training routines—these all are examples of passive experiments.

A torpid, inert, diffident, lifeless person is called passive with a tinge of deprecation. But the researcher is not always to blame for passiveness. An astronomer making his cometary observations can only passively registrate coordinates of a comet or vividly describe the picture. He can in no way exert any influence on its motion.

A processing engineer or operator manning a functioning installation can and must change the parameters of the process, but these changes are strictly prescribed by manuals and are quite insignificant. Many installations are fitted with an automatic controller, which maintains the parameters at a nearly constant level. Under these conditions the operator is a pretty passive figure, his functions being mostly reduced to logging.

Based on the findings of a passive experiment, we can derive the regression equation, and, as we found in the previous section, predict the response function from the regression equation within the region under study. Determination of values of the function within a region from separate known values at some points is called *interpolation*. To

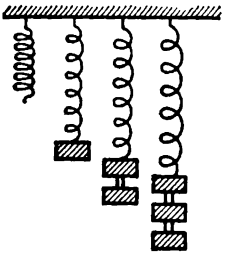


Fig. 23

be sure, you have encountered the term before. For example, in dealing with tables of logarithms or trigonometric functions, when it was necessary to work out the value of a function at a point not included in the table, we use linear or quadratic interpolation. But quite often it is necessary to know the behaviour of a response function beyond the region under consideration, i.e. to *extrapolate* the values. Could the regression equations be used to solve extrapolation problems?

In 1676 Robert Hooke published his law establishing a relation between the elongation of spring in tension and the acting (pulling) force involved.

Hooke's experimental arrangement was quite simple: to a suspended spring a force (a weight) is

applied and the elongation is measured (Fig. 23). Ever larger weights are applied and the results are put down. If in a plot the abscissa axis is elongation and the ordinate axis is weight, then the experimental points will thickly lie along a straight line (Fig. 24). This strongly suggests that the behaviour of the response function (elonga-

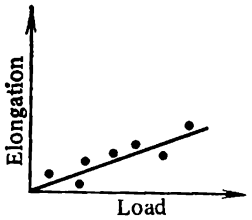


Fig. 24

tion-load dependence) can be interpolated between the points using a linear curve, i.e. we assume a linear dependence between the elongation and the load within the load range in question.

But can we predict the behaviour of the response function further, beyond the load range proved experimentally, by continuing the straight line? It is highly doubtful that we can. In fact, further experiments showed that beginning with some values of load the linearity is disturbed, and so does the elasticity. You will have got acquainted with the plot in Fig. 25 in your physics course, and it only serves here to illustrate the hazards of carefree extrapolation.

Recall the desalination section. Figure 12 is the plot of the relationship between salt concentrations at the input and output of the EDP within the interval from 100 to 1,000 milligrammes per

litre. Over the entire range the regression equation gives a straight line. But what is the behaviour below 100 milligrammes per litre? Can we make predictions within the range 0-100 milligrammes per litre by using the same equation, i.e. by

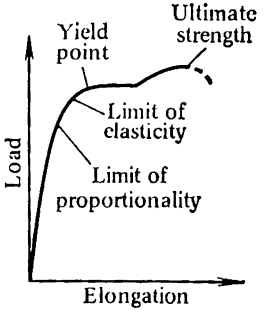


Fig. 25

continuing the line to the left till it cuts across the ordinate axis?

We have already discussed the question. We cannot, of course. Otherwise, it will give us 15 milligrammes per litre at the output with no salt at the input, a result without physical meaning. Clearly the curve must pass through the origin of coordinates, and so the curve will look like the dash line in Fig. 12. Consequently, here too the prediction of the input-output dependence appears erroneous, i.e. we cannot find the response function in a region where there are no experimental points using the regression equation derived from the results of a passive experiment.

Your refrigerator may be an object of some interesting research, if you care for it. You may

be interested in the variation of the temperature inside the refrigerator with the ambient temperature, the number of openings of its door, the amount of stuff in it, or any other factors subject to random variations. So you place a thermometer into the refrigerator and begin to put down its readings. Much to your surprise you will find that whatever the range of the variation of the parameters of interest (ambient temperature, amount of stuff, etc.) the temperature within the refrigerator—the response function here—only varies within a small interval from plus one to plus two degrees, and these variations are even difficult to detect with the help of your household thermometer. Thus, whatever the variations of the external factors, the variations of the response function are negligible and comparable with errors of measurement.

If now instead of your refrigerator you want to study a complicated process and you find the same picture, i.e. the output is essentially independent of the variation of input parameters, there is no way of constructing a reasonable mathematical model of the process based on the findings of a passive experiment. The situation generally implies that the process is so adjusted that it does not respond to permissible variations of input parameters. The experimental findings here appear to be concentrated within a small neighbourhood of one value of the response function, and we cannot solve any prediction problem, be it interpolation or extrapolation.

We will now look at some other aspects of the passive experiment, that set limits to its meaningful use. The factors may appear to be connect-

ed, or correlated, in some way or other. For example if we take a closer look at the sprinter's results, we will find a dependence between the duration of his training session and sleep he manages to grab, the amount of liquids consumed and the total amount of food, and so forth. This does not make our life easier. Remember that finding the coefficients of a regression equation by the method of least squares involves solving a system of linear algebraic equations. The coefficients of the latter are given in terms of the values of the factors, and their mutual dependence, deterministic or statistic, may make the matrix of the system ill-conditioned. The troubles involved in this case have already been touched upon in "Caution: the Problem Reduced to a Linear One"

Mutual dependence of estimates of the regression coefficients may also give rise to troubles. Consider this in more detail.

Suppose we have a simple regression equation with only two factors x_1 and x_2:

$$y = 0.2x_1 - 10x_2. \qquad (*)$$

Increasing the factor x_1 here increases y, and increasing x_2 decreases y. The sign at the coefficient of the factor thus points to the direction of variation of the response function with increasing factor. The value of the coefficients here is, of course, a measure of the rate of variation of the response function as an appropriate factor changes, and in equation (*) the factor x_2 is 50 times more "influential" than x_1.

In consequence, the absolute values of the coefficients indicate the relative contribution of the factor to the response function. And if, say, the

range of variation of the factors were the unit square $0 \leqslant x_1 \leqslant 1$, $0 \leqslant x_2 \leqslant 1$, then equation (*) could be simplified by discarding the first term, since its influence on the response function is negligible. Then

$$y = -10x_2.$$

The relative error here is not higher than 0.02.

Let us now write a similar equation in the general form

$$y = b_1x_1 + b_2x_2,$$

where b_1, b_2 are estimates (not exact values) of regression coefficients. If b_1 and b_2 are independent as random variables the picture will be the same: the magnitudes $| b_1 |$ and $| b_2 |$ will indicate the rate of variation of the response function with increasing x_1 and x_2, and their signs will indicate the direction of this variation. But if b_1 and b_2 are interrelated somehow, the picture is violated, i.e. uncertainties of determining one coefficient may influence the values of the other, even change its sign. The coefficients are now no longer a measure of the relative contribution of the factors and we cannot arbitrarily exclude from the equation the factors with "small" coefficients.

And the regression equation itself should not be scrapped. It may be of use in predicting the values of the response function within a region in question.

To sum up, passive experiment has many vices. To be sure, if there is no other way out, e.g. in cometary studies, a passive experiment may be a necessity. But in many problems in science and

technology we can forego the role of passive observer and pass on from passive experiment to active one.

The Active vs the Passive

We have already exposed the vices of passive experiment. But how can we eliminate them? We can do much to make experiment more effective.

Let us see what the experimentalist who wants to be more active has at his disposal, what he can change, select, discard, and what he should strive for.

Note that the primary objective of observations, experiments and trials is obtaining information about an object, process, or phenomenon. Accordingly, the active experimentalist's task is to acquire the desired information with the lowest costs and shortest time possible, or, if funds and time are short, his task is to accumulate as much information as possible given the constraints on funds and time.

But what information is he after? Natural phenomena are infinitely varied, objects of technology are involved, and there is, it seems, no useless, absolutely useless information. The philosophy behind the passive experiment is exactly like this: whatever we find out is good. The active experimentalist is no idler, he is keen on his dear problem and information he seeks is not just any data, but what will enable him to tackle his problem.

The active experimentalist is thus engaged in collection of information that, for one thing, is required to solve the problem at hand, and for the other, is sufficient for the purpose. If information is scanty, it should be obtained, and if it is impossible to obtain, then approximations will do, but incompleteness and limitations should be clearly perceived. Excessive information may be not only useless, but at times harmful, since it is not only a waste of time and funds, but may also be a spurious background against which useful information can be obliterated, or even it may give rise to prejudices.

The selection of information should thus be based on clear logical analysis of a problem. Speaking about information here we should understand it in a wide context. The very formulation of the problem, say, is also a piece of information about the problem.

The consistent analysis is by no means easy, not only the analysis of an experiment, but the very idea of such an analysis is hard to grasp.

Suppose a housewife wants to make a soup. What is she to strive at? Clearly, the soup should contain the appropriate amount of salt, vegetables should be boiled thoroughly but not cooked to a pulp, spices should be added in due time. But the soup may turn out to be a failure.

If you were asked to taste some kind of soup cooked by different housewives, would you be able to tell which is the best? Hardly, even if you would not be hampered by your desire not to hurt one of the women. And so instead of a clear answer you would mumble something incoherent. The difficulties involved have already been dis-

cussed in "A Glimpse of Criteria", and in the case of soup there is even more arbitrariness than in the essay issue.

In the case of a sophisticated piece of equipment, process or chemical product, the snags are essentially the same, or even more involved. Which of the characteristics of the process or product should be taken as outputs? Which of the measured or controlled variables should be taken as inputs (in the theory of experiment they are called factors), the remaining ones being dumped under the heading of uncontrolled variables? We have already discussed the problems involved in ore refining, where there are about two hundred controlled, measured or monitored parameters. Unfortunately, the situation is not that rare.

To be more specific, we will take some practical problem. In recent years my colleagues and I have been dealing with it. It is the eternal problem of lubricants, most of which are derived from petroleum.

Normally they remove undesirable components from these oils. But even the most sophisticated procedures fail to give oils with perfect characteristics required by modern industries. So oils must be stable to oxidation, noncorrosive for metal surfaces, they must reduce wear of friction surfaces, and must have good washing properties. These properties are generally achieved using special-purpose chemical compounds, called additives.

Small amounts of additives, from fraction to several per cent, drastically improve the quality of oils. It would seem that the present state of

the art in chemical sciences would make it possible to describe the action of additives, to select the best additives and their optimal percentages. But such a theory is nonexistent so far. Therefore, the set of additives and their concentrations are selected experimentally.

Let us now concentrate on the optimal ratio of additive concentrations. Schematically, the

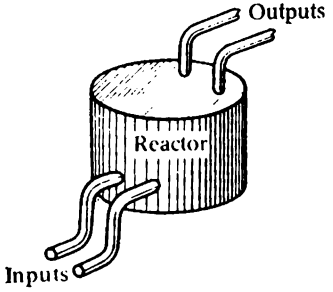


Fig. 26

situation may be represented in the following way. There is a vessel, or reactor, partly filled with the oil base, i.e. semifinished oil to be improved upon and optimized. The vessel has several inputs (tubes with cocks) used to feed one of the additives (Fig. 26). Of course, effective stirring is provided in the vessel and some of the liquid is taken away for analysis through the tubes at the top, the outputs. We will assume that at each of the outputs some definite parameter is measured. This scheme may be a model of any other process: the tubes are inputs, the cocks are controllers, and other tubes are outputs.

In this problem we shall achieve optimal ratio of additive concentrations. But what is to be understood under the optimal ratio? Process engineers attach importance to a number of parameters. Acid number, say, characterizes stability of an oil to oxidation by oxygen in the air, and so it should be made as low as possible. Corrosiveness, too, should be reduced; it is characterized by the percentage of metal corroded away in a liquid. The corrosive effect of the liquid on different metals is different. Which metal then must be taken as a reference?

Oil is known to stiffen with time—its viscosity increases—thus substantially impairing the functioning of the oil. We should therefore try and maintain the viscosity at the same level.

Also rubber elements of apparatus swell in contact with oil, and so this swelling must be as small as possible.

The examples of these parameters could be multiplied.

An important fact was found experimentally: if several additives are combined, their effect can sometimes be higher than the sum of individual effects. This effect was called *synergism.*

Synergism may enable, for example, the resistance of oil to oxidation to be improved or the consumption of additives to be reduced. In certain concentrations, however, additives may appear to be antagonists. The qualitative parameters may be interrelated somehow. What is more, improving one of them may impair another one. Accordingly, it is impossible to achieve optimization of all of them. The situation is not

new, we have discussed it in detail in "A Glimpse of Criteria"

Commonly, the situation is like this: customers come up with their requirements, which are at times contradictory, and it may be either difficult or impossible to meet them accurately. And so the designer will have to work out some compromise.

In this problem we could select some economic criterion, for example, reduced costs to be minimized, but now designers follow another procedure. They take as the criterion a figure of merit of the oil that is the most critical in this situation, e.g. showing the largest spread in manufacture or not meeting the specification, or something of the kind.

For the first time we encountered the problem of selecting the ratio of additive concentrations, when the designers were asked to minimize the acid number. And so it was selected as the figure of merit, or criterion of quality. On all the other output characteristics they only imposed some constraints, i.e. some limits were set within which the parameter was allowed to vary.

We are now able to formulate the optimization problem: to work out the additive concentration ratio such that the acid number be minimal, given the specified constraints on the remaining output variables. Later we had to change the criterion and optimize another output parameter, but the problem remained essentially the same.

For the problem to be solved, the designer must acquire a thorough knowledge of the plant, process, and mechanism. Which part of the knowledge should be used? The question is not easy to

answer: we are in possession of knowledge whose value and reliability are fairly different. What literature evidence is to be relied on, and what not? What are we to make of the data obtained for a similar process but on another installation, differing in construction, using other raw materials, functioning under other working conditions? Or maybe we should rely on our own experience, although it comes from disparate objects, just because the experience is ours, i.e. it has repeatedly proved its value.

But suppose we have already chosen the inputs and outputs, have our criterion of quality and must now set out to perform experiments. How many experiments are to be carried out? At what values of input factors, i.e. at what points within the permissible region of the factor space should the experiments be done? And in what succession should the points be covered?

Let us begin with the first question. We will have to take a close look at the region of possible variation of factors, for which purpose we will have to vary individually each of the factors, fairly often in the interval of its variation, the others being fixed, and thus to exhaust all the possibilities. Granted, we will thus find the optimum. But just imagine how many experiments we will have to carry out.

What is important here is the number of distinguishable values. If factor x varies from 0.1 to 3.3 and is measured to within 0.2, then it will have

$$\frac{3.3-0.1}{0.2} = 16 \text{ distinguishable values.}$$

Clearly, we simply do not need more precise measurements. The values of factors to be varied in the theory of experiment are called *levels*. The number of levels for all the factors determines the number of the values of inputs. If we start with five additives and each is varied through five levels, then the total number of states will be 5^5 3,125.

Experimental determination of the oil characteristics studied may take two weeks, and to cover 3,125 points of the factor space will take more than 12 years, a time during which not only some of the experimentalists will retire, but also the oil specifications will change—a natural consequence of the vehement development of technology. We are thus not satisfied with an exhaustion of all the values possible. What is to be done then? We cannot possibly forego a possibility of improving upon the oil just because we cannot go through the entire exhaustion of all the combinations.

The situation seems to be an impasse: either to perform the many experiments taking months and years, or to perform a small number of experiments and select the best one, ignoring the real possibility of finding a combination of factors, in which the figure of merit will be much higher than the one found randomly. Of course, we can always justify the situation by saying that we have anyway found a combination of factors much better than the one used previously.

With such a trial-and-error method the probability of chancing upon an optimum is next to nothing. It is here that the logical analysis of experiment comes in. The latter sections will be

devoted to the analysis of the possibilities of the experimentalist and to experiment design, which is not only much more effective, but leads to success.

Who Has the Best View?

In an active experiment we are able to select in the factor space those points at which we want the experiment to be done. To be sure, this is a great advantage over the passive experiment, where we are at the mercy of chance.

But reasonable, purposeful selection of points in factor space is not a simple problem, and most of the publications on experiment design discuss the selection of points in a factor space.

The points selected and the sequence of experiments at the points selected are referred to as the plan of experiment, and the selection of points and strategy to be followed is called the *experimental design.*

To be more specific, let us consider a practical example of automatic correction of concrete mix composition, a problem of importance for huge dam projects.

Concrete should be uniform. Cement is more expensive than aggregate and so the proportion of cement should be reduced as much as possible. In hydropower dam projects concrete work accounts for the major share of the effort, which makes automatic correction of the concrete mix a must and justifies the sophisticated and expensive control equipment used for the purpose.

Before we set out to design the correction sys-

tem, we should take into consideration the following two points. First, the behaviour of a concrete and concrete mix is critically dependent on the humidity of aggregate, for the most part sand. Second, of no less importance is the granulometric composition of aggregate (sand, gravel, crushed rock, slag), since in the final analysis these determine the water requirement of a given mix, i.e. water content at a given consistency.

Systems have been developed which correct the water content depending on the humidity of sand. Such systems may only be effective with highly uniform properties of the aggregate or relatively slow rate of their variation and small fluctuations of the humidity of the coarsest component.

But the required uniformity of aggregate properties can only be provided at small concrete-mixing plants. On a huge dam project, however, it is virtually impossible to provide such a homogeneity.

Some observations show, for example, that the humidity and granulometric composition of aggregate, especially sand, vary at a high rate. Under these conditions correction systems will be advisable that follow the fluctuations of humidity and granulometric composition of the aggregate.

The problem thus reduces to the development of an active system that would automatically determine optimal water requirement for the mix.

Clearly, to each combination of aggregate (granulometric composition) there corresponds some minimal water requirement, the domi-

nating factor here being the sand content in the aggregate.

The dam engineers know from experience the dependence between the true water requirement and the sand proportion in the aggregate combination—it is close to parabolic curve. The extremum of the parabola corresponds to the minimal water content. If we denote by v the water requirement and by x the sand proportion, then they will be related by the empirical relationship

$$v = b\,(x + a)^2 + c, \qquad (*)$$

where $b > 0$, c and a are unknown constants, which define the shape and position of the parabola. As the granulometric composition varies, so do the position and shape of the parabola, and point a—the coordinate of the extremum—shifts.

The problem thus boils down to maintaining the consistency of the mix within preset limits and to finding the optimal proportion of sand at which the water requirement is minimal. The permissible percentage of sand x varies within the limits $x_{\min} \leqslant x \leqslant x_{\max}$. The unknown parameter a can be found simply by specifying three values x_1, x_2, and x_3 of the variable x within the interval $(x_{\min}, x_{\max})$ and finding experimentally the appropriate values v_1, v_2, and v_3 of the function v. The values of a, b, and c will then be found from

$$\left.\begin{aligned} v_1 &= b\,(x_1 + a)^2 + c, \\ v_2 &= b\,(x_2 + a)^2 + c, \\ v_3 &= b\,(x_3 + a)^2 + c. \end{aligned}\right\} \qquad (**)$$

Solving these equations is child's play, but then some complications await us.

In solving the equations it is immaterial what values x_1, x_2, and x_3 are chosen within the permissible interval $(x_{\min}, x_{\max})$. But v_1, v_2, and v_3 are determined experimentally, and hence it involves some uncertainty or error, which, when taken into account, may drastically change the picture.

Let us now simplify the problem a bit, by putting $b = 1$, $c = 0$, so that the equation takes the form

$$v = (x + a)^2, \qquad (***)$$

and let the independent variable x vary within the limits from $x_{\min} = -1$ to $x_{\max} = +1$. This can be always achieved by simply changing the variables.

Thus, it is required to find the position of the extremum of the parabola (***), given that x varies within the limits $-1 \leqslant x \leqslant +1$. To this end, we will now simply have to find two values v_1 and v_2 at some x_1 and x_2, and it is clear that these values of x can be chosen arbitrarily within the interval $(-1, +1)$. But what values of x_1 are to be chosen, if v is measured with an error? What is more, now the question arises of the necessary number of points x_i where measurements are carried out and the number of measurements at each of them. At this point we have to restate the problem.

We will assume that $v_i = (x_i + a)^2$ are measured with a random additive error ε_i, i.e. measurements give

$$y_i = v_i + \varepsilon_i.$$

The errors in finding v_i lead to errors in finding a, and it is only natural to take up the problem of experimental design, where a will be found with the least possible error. Here, just as in the problem of choosing the optimal strategy of car insurance discussed earlier in the book, it would be desirable to exclude the dependence of the error on a. To this end, its a priori distribution is introduced, and the error is averaged once more over this distribution.

We will not discuss the formalism involved, since it is fairly complex. Instead, we will try to approach the estimating a from a simpler and more graphic angle.

It would be reasonable to assume that errors ε_i are independent both of x_i and of time. How then are we to select the values of the independent variable x_i that provide a high accuracy of finding a? Since ε_i are independent of x_i, the relative error varies with v_i and it is thus clear that at different x_i the magnitude of the error will be different. Accordingly, it would be quite reasonable to state the problem as selection of the values of x in the permissible interval, such that the respective errors be minimal.

We do not know where a lies within the interval $(-1, +1)$, and the parabola looks like that in Fig. 27. Clearly, the function $v(x) = (x + a)^2$ attains its maximum at one of the ends of the interval $(-1, +1)$, i.e. either at $+1$ or at -1, but which one we do not know. Therefore, we must use both extreme values.

If we make only two measurements, one at each of the points $x = -1$ and $x = +1$, we

will have

$$y_1 = (-1 + a)^2, \quad y_2 = (1 + a)^2.$$

After some algebra we get $a = (y_1 - y_2)/4$.

Measurements can be made repeatedly, however. The same arguments lead us to the conclusion that it would be expedient to perform measurements at the extreme points -1 and $+1$

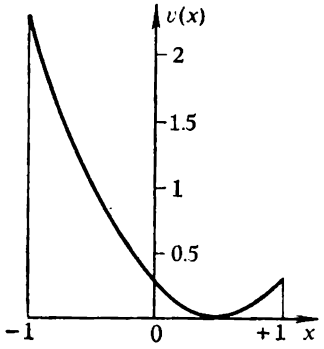


Fig. 27

alone, since measurements at internal points of the interval $(-1, +1)$ may yield relative errors larger than that for the largest of $v(-1)$ or $v(+1)$. The reasoning seems to be plausible, although we have not proved it. In actuality, it is wrong: the optimal strategy appears to be quite different. However, the measurement strategy in which a half of measurements are carried out at -1, and the other half at $+1$ (we will call it suboptimal strategy) appears to be fairly good. We have already discussed the optimal

strategy: finding x_i where measurements are to be performed requires a difficult problem of variational calculus to be solved. Comparison of the suboptimal strategy with the optimal one shows that the suboptimal strategy used at $2n$ measurements (n measurements at each of the extreme points of the interval $(-1, +1)$) gives smaller root-mean-square errors than for n measurements in the optimal strategy. This fact confirms at once two intuitively immediate points: first, selection of x_i at which measurements are to be made is important and difficult; second, in the parabola example the extreme values -1 and $+1$ are significant.

It is exactly upon this suboptimal strategy that the mix correction system was predicated.

To sum up: in experimental design or analysis one should give careful thought to the values of an independent variable at which the experiment is to be performed. Admittedly, the above example does not enable the infinite variety of situations possible to be predicted, and so we will return to the question of where experiments are to be done later in the book.

Step by Step

A team of hikers is going to climb a mountain. You make yourself comfortable in a chair near the mountain and watch their progress through the binoculars. Using some device they measure the height over sea level. After they have made 500 paces due east they measure the

height again and now make their way due north, due west, and due south. At each of the predetermined points they repeat the procedure, measuring in succession the height at each of the points of a network with the spacing between the points 500 paces.

Their objective was to reach the summit, i.e. the highest point in the neighbourhood.

Fortunately real hikers do not follow this procedure, although experimentalists do. Why such senseless waste of time?

Admittedly, it is possible to work out the highest point on a surface by systematically measuring the heights of the points lying on the surface and selecting the highest. But the reasoning is far from the only one and it by no means follows in a natural way from the very formulation of the problem. Just the contrary is true, it is a bad and unnatural way of looking at things.

Hikers generally make the following: they survey the neighbourhood, choose the most convenient path to the summit and do not care to explore all the points within some network.

Experimentalists sometimes rely on exhaustion of points, either on a network or other array, because they fail to give a clear formulation of the problem at hand. One of the achievements of the theory of experiment is the development of a clear, consistent analysis of the process of putting forward hypotheses, of formulation of problems and hypothesis testing.

Let us now return to the oil additive problem. Remember the important thesis: the success of statistical analysis is determined by the rigour of the formulation of the problem.

In the problem we have to find the optimal point, i.e. a ratio of concentrations of additives that will yield the best result. We will have to work out a rule to be followed in performing the experiments, which would enable us to find the optimal ratio of the additives for a minimal number of experiments.

The shortest, if not the most convenient, way to the summit is to leave each point along the steepest slope. If there is only one summit without smaller, or local, ones the procedure will lead you to success whatever point on the slope you start from.

And still there is a difference between the hiker and the additive expert. The latter does not see his "mountain"—the response surface, which makes it seem that he is in absolutely different situation. This is not so, however.

In fact, let us equalize the conditions for the hiker and the chemist by suggesting the hiker to climb the mountain on a moonless night. His strategy will be quite obvious: he will make small steps to the right and left, forwards and backwards, and move in the direction in which the slope is the steepest. In this way he will reach the peak. To acknowledge the fact of reaching the peak is no problem as well: if steps in the four directions bring him down a bit, it implies that the hiker is at the summit.

Our chemist should follow the same way in seeking the optimal concentrations of additives, and so he should start with experimental design.

Recall that we have selected acid number as our optimization parameter, the optimal compo-

sition being one for which the total acid number is minimal. Accordingly, the independent variables, or factors, are the very concentrations of additives, and the output, or response function, is the total acid number.

In these terms, the problem is to seek not the maximum (a peak on the surface), but a minimum (the lowest point). The procedure of looking for a minimum or maximum is absolutely the same: an ant travelling from the edge of the brim of a hat does not care if the hat lies with its brim up or down (Fig. 28).

If now we make no assumptions as to the structure of the response surface, our situation will be hopeless. Suppose we have only two factors, so that the response surface is a conventional surface. For example, if we do not assume that the surface is continuous, it may behave in an irregular way, showing no minima or maxima in a conventional, graphic way If the surface is sufficiently smooth, without any abrupt walls, it may look like a bunch of grapes (Fig. 29), with its multitude of minima (mathematicians call them local), and once you have got in one of them (say, the one marked in the figure) it is by no means easy to figure out if there are any other minima, which are yet lower. It is all the more difficult to achieve the lowest point, the global minimum.

It is hard to believe that nature is so insidious. It is only natural, at least at first, to suppose that the response surface is structured in a simpler way, without such a multiextremality, although one cannot a priori exclude a possibility of any local extrema.

Let us mark some point on the surface and look at its neighbourhood. In the immediate vicinity a piece of sufficiently smooth surface is virtually undistinguishable from a patch of plane, and if

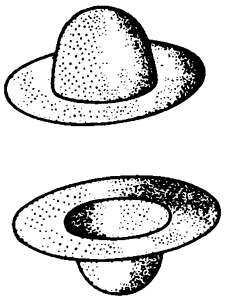


Fig. 28

the plane is not parallel to the horizontal plane, then we can make a step down the steepest slope. We thus go over to a new point and again mark off its immediate vicinity, construct a patch of plane, which is virtually undistinguishable from the piece of the surface, and make a further step down the steepest slope.

We will carry on the procedure until we reach the boundary of the permissible region or the plane in the vicinity of the final point will appear to be parallel to the horizontal surface. This point, as you remember, is called stationary. If we place a heavy ball at a stationary point, the ball will remain at equilibrium there. Besides equilibrium points there are quasi-equilibria—saddles or cylindrical surfaces with the generatrix parallel to the horizontal plane.

To analyse the behaviour of a surface in the vicinity of a stationary, but "suspicious", point, the surface should be approximated by a second-order surface. After its features have been found,

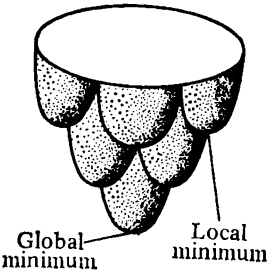


Fig. 29

we can heave a sigh of relief, if we have arrived at a minimum. But even now we are not on a safe ground, as we seek the global minimum, and we should see to it that we are not at a local one.

So you see that the above step-by-step, or sequential, strategy is a form of the Wald sequential analysis, in which the hypothesis (about the point being stationary, about a minimum or a global minimum) is tested by experiments. We will not here dwell on the details of search of extrema.

Where Are Experiments to Be Staged?

In the previous section we looked at the situation with one factor and parabolic response function. And now we will turn to multifactor problems.

To begin with, consider the simple problem of weighing any three objects A, B, and C. The first idea to occur will be to weigh each of the objects in succession. This is exactly the procedure followed by the traditional natural scientist, but at first he makes an "empty" weighing to determine the null of the balance. When an object is placed on the balance an entry $+1$ is made into a table, when it is absent on the balance, -1 is entered. The results will be denoted by y with an appropriate subscript (see Table 4).

Table 4

Traditional Weighing Procedure for Three Objects

Trial run	A	B	C	Results
1	-1	-1	-1	y_0
2	$+1$	-1	-1	y_1
3	-1	$+1$	-1	y_2
4	-1	-1	$+1$	y_3

The experimentalist here studies the behaviour of each factor separately, i.e. performs unifactor experiments. The weight of each of the objects is only estimated from the results of two trials, one of them being the "empty" one, the other the trial in which the object was on the balance. For example, the weight of object A is

$$A = y_1 - y_0.$$

Generally, the weighing error is assumed to be independent of the object being weighed, the weight being an additive quantity having the same distribution. The variance of the weight

of an object will be

$$D(A) = D(y_1 - y_0) = Dy_1 + Dy_0 = 2\sigma^2,$$

where σ^2 is the variance of any weighing. The variance for B and C will be the same.

But the experiments can be done using another plan, a multifactor one. It is illustrated in Table 5.

Table 5

Multifactor Plan of Weighing Three Objects

Trial run	A	B	C	Results
1	+1	−1	−1	y_1
2	−1	+1	−1	y_2
3	−1	−1	+1	y_3
4	+1	+1	+1	y_4

Now we have no "empty" weighing. In the first three trials objects A, B, and C were weighed in succession, and in the fourth one all the three objects together were weighed.

By multiplying the elements of the last column of the table one after another by the elements of columns A, B, and C and dividing by two because, according to the plan, each of the objects has been weighed twice, we obtain

$$A = \frac{1}{2}(y_1 - y_2 - y_3 + y_4),$$

$$B = \frac{1}{2}(-y_1 + y_2 - y_3 + y_4),$$

$$C = \frac{1}{2}(-y_1 - y_2 + y_3 + y_4).$$

Now the weights of the objects are not distorted by other weights because, say, the expression for B contains each of A and C twice and with different signs. The variance of the weighing error will now be

$$D(A) = D\left(\frac{y_1 - y_2 - y_3 + y_4}{2}\right) = \frac{1}{4} 4\sigma^2 = \sigma^2,$$

i.e. half that of the variance in the unifactor experiment. If we wanted with the unifactor plan to obtain the same variance as with the multifactor plan, we would have to perform each of the four unifactor trials twice, i.e. to carry out eight weighings instead of four.

Consequently, using the multifactor plan each weight is derived from the results of all the four trials, which accounts for the halving of the variance.

Exactly the same situation will occur in dealing with a response function that linearly varies with three factors x_1, x_2, x_3.

Remember the desalination process where the amount of salts at the output of the plant (y) varies with the amount of salts at the input (x_1), the amount of demulsifier added (x_2), and the residence time in the electric field (x_3). When these factors vary within some limits, y will vary linearly with x_1, x_2, and x_3.

The regression equation here will be

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3. \qquad (*)$$

We will vary each of the factors at two levels, taking the levels to be the largest and smallest values of the factor with the interval of its variation and assigning to these levels the symbols

+1 and −1. However, as was pointed out in the previous section, some linear substitutions of variables make it possible for the factor to vary within the interval (−1, +1).

We can now make use of the design matrix given in Table 5 to carry out our experiments. We will only present it in new variables and add a column of another imaginary variable, x_0, required to estimate the free term β_0.

According to the plan, trials are performed at the following points of the factor space: in the first trial x_1 is at the upper level, and x_2 and x_3 at the lower level, i.e. in the transformed variables the experiment is done at the point (+1, −1, −1); in the second trial, x_2, is at the upper level, and x_1 and x_3 at the lower level, i.e. at the point (− 1, + 1, − 1), and similarly, in the third trial, at the point (− 1, − 1, + 1), and in the fourth trial, at the point (+ 1, + 1, +1).

After the plan has been realized, four equations in four unknowns are obtained. The solutions of them will be the estimates of all the four regression coefficients β_0, β_1, β_2, β_3. In the plan of Table 6 the number of trials is thus equal to the number of constants to be determined. Such plans are called saturated.

Note that we have used not all the points with "extreme" coordinates, i.e. $\pm$ 1, or put another way, not all the combinations possible. Indeed, all the possible combinations of three symbols, each of which takes on the values either + 1 or − 1 will be $2^3 = 8$. We have only used 4 of them so far. How about the remaining ones?

In order to be able to answer this question, let us turn to a simpler situation, where we have

Table 6

Design Matrix for a Linear Model with Three Independent Variables

Trial run	Plan				Results
	x_0	x_1	x_2	x_3	
1	+1	+1	−1	−1	y_1
2	+1	−1	+1	−1	y_2
3	+1	−1	−1	+1	y_3
4	+1	+1	+1	+1	y_4

only two factors, and only two levels of them. A plan given by all the possible combinations of the two levels (it is called the complete factor plan) will contain $2^2 = 4$ points, they are represented in Table 7 by two middle columns.

Table 7

Design Matrix for the Complete Factor Experiment of the Type 2^2

Plan			
x_0	x_1	x_2	x_1x_2
+1	−1	−1	+1
+1	+1	−1	−1
+1	−1	+1	−1
+1	+1	+1	+1

If we now carry out experiments according to such a complete factor experiment, we can estimate all the coefficients in the regression equa-

tion

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad \beta_{12} x_1 x_2.$$

The last term here is no longer linear. It contains the product of factors and therefore it is called the interaction effect, although the interaction may be much more complicated. But such is the currently adopted terminology. The complete factor experiment thus enables us to estimate the coefficients of a more general equation than the linear equation in two variables.

When there are serious grounds to believe that $\beta_{12} = 0$, then in the matrix of Table 7 we can put $x_1 x_2 = x_3$, and obtain the matrix of Table 6, i.e. a plan for a three-dimensional factor space, although now it is not a complete factor plan for three variables, but its part. Such an experiment is called a *fractional factorial experiment*, and its plan a *fractional replication*, so that Table 6 is a fractional replication of the complete factorial plan of type 2^3, whose matrix is represented as the second, third and fourth columns of Table 8.

It is worth noting that the design matrix for a linear model in three variables given in Table 6 is part of the matrix of the last plan—it consists of the four lines of the first four columns and is enclosed in a dash line box. Therefore, the plan in Table 6 is called one-half replication of the complete factorial experiment and is denoted 2^{3-1}. If we change all the signs in this one-half replication, we will get the lower four lines of the same matrix, i.e. the other one-half replication.

The beginnings of experimental design date back to the 1920s, when the English statistician

Table 8

Design Matrix for the Complete Factorial Plan of Type 2^3

Trial run	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+1	+1	−1	−1	−1	−1	+1	+1
2	+1	−1	+1	−1	−1	+1	−1	+1
3	+1	−1	−1	+1	+1	−1	−1	+1
4	+1	+1	+1	+1	+1	+1	+1	+1
5	+1	−1	+1	+1	−1	−1	+1	−1
6	+1	+1	−1	+1	−1	+1	−1	−1
7	+1	+1	+1	−1	+1	−1	−1	−1
8	+1	−1	−1	−1	+1	+1	+1	−1

Sir Ronald Fisher published his first works on the subject. His ideas were developed in the 1950s by the American mathematician J. Box and his co-workers, and it was these works, which were clearly applied in their nature, that contributed to the wide recognition of the theory. But the terminology of Box does not appear convenient, because many known concepts for which there are already established terms in control theory or statistics were called differently.

The complete 2^3 factorial design (Table 8) enables us to estimate the coefficients of the regression equation that contains three pairwise interactions and one triple interaction. The respective products are given in the upper line of Table 8, and so you may get some idea of the form of the regression equation. If the experimentalist is confident that the response surface is linear, i.e.

that there are no nonlinear terms of the regression equation, then he can introduce the new variables $x_4 = x_1x_2$, $x_5 = x_1x_3$, $x_6 = x_2x_3$, and $x_7 = x_1x_2x_3$, and obtain a design matrix to estimate the eight coefficients (including β_0) in the linear regression equation with seven factors.

If the problem at hand allows a linear approximation, then in the complete factorial experiment there will be many "extraneous" trials. So with three factors, as we have seen, we can compute the regression coefficients in the linear equation with only four trials, and in a 2^3 complete factorial experiments we have eight and so four of them are "extraneous". The results of these trials can be used in two ways: first, to get more accurate estimates of regression coefficients, second, to test the adequacy of the model constructed. But with seven factors the complete factorial experiment at two levels contains $2^7 = 128$ trials, and, as it was just mentioned, it takes only eight trials to work out eight coefficients of the linear regression equation. We thus end up with 120 "extraneous" trials, and it is by no means necessary to realize all of them. It only suffices to use some of them to test the adequacy and refine the estimates.

We can carry on reasoning along these lines, but it appears that the general procedure is clear. Really, there are an infinite variety of plans, different approaches and possibilities of reducing the number of trials necessary to arrive at more complete and reliable information. But this book is not a text in experimental design, and so the reader is referred to the special literature on the subject for details.

It is only important here that the reader perceive the advantages of the designs just described. For example, all the coefficients in the regression equation are estimated independently. This implies that the coefficients, say, in equation (*) indicate the relative contribution of appropriate terms, and hence we can ignore terms negligible as compared with the others. In other words, factors with relatively small coefficients can be discarded as insignificant, without recalculating the coefficients.

If the response surface in the vicinity under consideration is nonlinear, then two-level design will be insufficient, and so we will have to use three levels. In addition, we will have to increase the minimal quantity of experiments. By way of illustration, we can return to the additives for oils.

The preliminary analysis indicated that the criterion of quality—the acid number—varies linearly with the concentrations of additives, and so the response surface may be rather complex here.

At the same time, only two of the five additives noticeably changed the acid number in varying the concentrations, whereas the remaining ones exerted nearly no influence on the criterion chosen. Therefore, the problem reduced to a two-factor design, and it can be conveniently used to give a graphic illustration of the successive situations in search for optimal concentrations. We will refer to the two additives as D and E.

As a first step of the design, the concentration range from 0 to 1.4 per cent was chosen, based on the experience available. The surface in this

region was assumed to be nonlinear, and the simplest surface here will be the second-order one. The regression equation will be

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \\ + \beta_{22} x_2^2.$$

It will be recalled that second-order surfaces are classified into ellipsoids, paraboloids, hyperboloids, and cylinders depending on the signs of β_{11} and β_{22}, and the sign of the discriminant $4\beta_{11}\beta_{22} - \beta_{12}^2$ of the quadratic form, i.e. the form of the surface is determined by the magnitudes and signs of the three last coefficients.

The coefficients were looked for, using a three-level design. The levels were the extreme values within the intervals of the variables (after a transformation they, as before, have the values ± 1), and the middle point, which transforms into the point with a zero coordinate.

Table 9 gives the conditions and results of trials according to a second-order design with two variables. After the results have been processed and insignificant coefficients discarded, the resultant model of the response surface took the form

$$y = 0.78 - 0.069x_1^2 + 0.158x_2^2.$$

This is the surface of a hyperbolic paraboloid shown in Fig. 30. Motion along the D axis reduces y, and so we will have to move in that direction. Further steps brought us to point q corresponding to the concentrations 0.45 per cent (F) and 5.5 per cent (D), in the vicinity of this point the response surface given in Fig. 31 has a distinct minimum.

Table 9

Conditions and Results of Lubricant Oxidation Experiments for D and E Additives
(x_1 is the concentration of D; x_2 is the concentration of E)

Trial run	x_1	x_2	x_1x_2	x_1^2	x_2^2	D, %	E, %	y
1	+1	−1	−1	1	1	1.207	0.207	0.99
2	+1	+1	+1	1	1	1.207	1.207	0.76
3	−1	+1	−1	1	1	0.207	1.207	0.80
4	−1	−1	+1	1	1	0.207	0.207	0.76
5	−1.414	0	0	2	0	0	0.707	0.73
6	0	+1.414	0	0	2	0.707	1.414	1.14
7	+1.414	0	0	2	0	1.414	0.707	0.60
8	0	−1.414	0	0	2	0.707	0	1.10
9	0	0	0	0	0	0.707	0.707	0.83
10	0	0	0	0	0	0.707	0.707	0.78
11	0	0	0	0	0	0.707	0.707	0.72
12	0	0	0	0	0	0.707	0.707	0.85
13	0	0	0	0	0	0.707	0.707	0.74

The equation of the surface will now be

$$y = 0.148 - 0.052x_2 + 0.093x_1^2 + 0.073x_2^2,$$

so that the minimum is achieved at a point corresponding to 0.54 per cent of E and 5.5 per cent of D. The acid number y is here 0.14, which is much less than any of the results obtained in the first region selected by the experimentalists (the last column in Table 9), where the minimal value of the acid number was 0.6.

When another pair of additives, we will call them F and G, are used, the response surface takes the form as represented in Fig. 32. This is also the hyperbolic paraboloid, but in this case the surface touches the plane $y = 0$, and hence

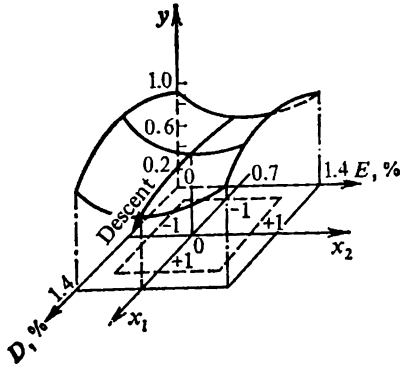

y
1.0
0.6
0.2
0.7
1.4 E, %
Descent
-1
0
+1
x2
x1
D, %
1.4

Fig. 30

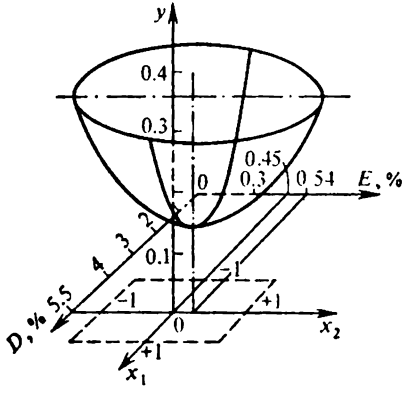

y
0.4
0.3
0.45
0.3
0 54 E, %
0
0.1
4
3
2
D, % 5.5
-1
+1
0
x2
x1

Fig. 31

the optimal points lie near the level $y = 0$ on this surface, e.g. point q_1. But you should not think that you can really obtain a zero acid number. You should take into account the experimental errors and remember that all the quantities are given here to within this error.

We will not consider other examples. It is hoped that you already understood that there is

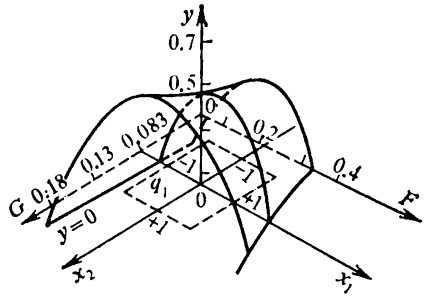


Fig. 32

a rich variety of forms of response surfaces in the vicinity of the optimal point, and their study and interpretation in the language of the field take some grounding in the trade.

Thus, to find the optimal combination of additives we used the step-by-step strategy of motion over the response surface along the steepest descent direction in each subregion. But if we sought for a maximum, not minimum, we would have followed the steepest ascent direction.

To sum up: in a step-by-step strategy we at first explore a small part of the response surface, construct a model for this part of the surface, test the hypothesis that the optimum has been achieved

and make one of the decisions: YES, NO or MAYBE. If the decision is YES the search is discontinued, if NO, we make one more step in the steepest descent (ascent) direction, and go over the procedure again. If then the decision is MAYBE, more experiments are required to get a better idea of the form of the response surface. This is an application of the Wald sequential analysis in experiment, a major breakthrough in the theory of experimentation.

However, in experimental design the sequential strategy is not the only achievement. The multifactorial experiment, i.e. rejection of traditional variation of factors one at a time, the remaining one being fixed, turned out to be a no less remarkable breakthrough than the sequential experiment. These strategies markedly reduce the number of experiments required. So if it is necessary to carry out an experiment with four factors and five levels, the total number of trials will be $5^4 = 625$. If we apply one of the forms of optimization of experiment (saturated D—optimal design), the same results can be obtained after 15 trials.

Coming under this heading is also such an order of experimentation, which does not lead to some prejudices and kills any systematic uncertainties, which are extremely difficult to get rid of in a passive experiment. Here the statistician, paradoxical as it may appear, is helped by chance. Later in the book we will consider this in more detail.

Ways to Success

Among the 1,093 patents granted by the US Patent Bureau to the famous Thomas Alva Edison was the patent No. 223898 as of 27 January 1880 on the carbon filament lamp. Three thousand people came over in special-purpose trains ordered by Edison to take a look at hundreds of electric bulbs hanging over in his laboratory and nearby roads in Menlo Park, New Jersey. But before this triumphant demonstration Edison screened six thousand carbon-containing substances from conventional sowing cotton covered with carbon to food-stuffs and resins. The best candidate turned out to be bamboo of which the case of a Japanese palm fan was made.

You understand, I think, that to try six thousand filaments took tens of thousands of trials, this gargantuan effort took about two years. Obviously, if Edison knew the theory of experiment, the number of his trials could be drastically reduced, maybe several-fold. But at the time of Edison the factor analysis was not yet available, besides Edison was suspicious of statistics, a science that did not agree with his education and temperament.

We could look at the work of Edison from modern viewpoint, but we will rather take a simpler example from real life.

Everything in the life of a singer depends on her success, especially at contests where she needs not just success, but a triumph. And suppose she seeks advice from her mathematician friend about an aria and attire to be selected. Just imagine:

an egghead and singing! But instead of admiring her stage dress or her vocal faculties he suggests to explore the response of the public to different attires and pieces of singing. Within the time space before the contest she can test herself in the public purposefully. For this purpose her mathematician friend suggests a method similar to the one used in the lubricant additive problem.

Unlike the situation with the additives, factors here are not quantitative but qualitative. The singer has three dresses and five specially prepared arias. There is not any numerical variable varying continuously from the black silk dress to the folk costume. By the way, the continuity of passing over from one of the dresses to another is irrelevant here, since we can assign number to the factors, quantitative or qualitative, and to versions, to be called levels here. But when a factor is quantitative, just like concentration or weight, the order corresponds to increase or decrease in the value of the level. If the factor is qualitative, then the order of numeration is immaterial. Therefore, qualitative factors call for another approach.

Some thought should also be given to the results of observations, i.e. to a way of measuring success, or comparing successes of different performances.

To characterize success we can rely on the duration of applause or its intensity, or else the number of encores. Also we can think of some comparative characteristics, say, to nominate three categories: great success, medium success, and small success, and so on. To be sure, when looking for some subjective criterion we may face with new problems: who is to be taken as an ex-

pert, should it be always the same person, could the singer herself be trusted to make estimates? We are not going to dwell on this subject here, we will just suppose that there is some criterion.

Besides, the response is conditioned by the audience: a theatrical crowd, the participants at some scientific conference, or night-club haunters—the reaction will be predictably different. The situation thus boils down to the three factors: arias, dresses, and audiences. We will have five arias (1, 2, 3, 4, and 5), three dresses (A, B, and C), and five audiences (α, β, γ, δ, and ε).

Each of the factors contributes to the parameter we use to measure success, and each factor assumes different levels. It is precisely because of the inhomogeneity that we have to seek the best combination.

A model of dependence of the parameter measured (y) on the factors can be written as in the regression model in the form of the sum of effects of each factor and interaction effects. By way of illustration, we will write a model for two variables—arias and dresses—using conventional denominations

$$y_{ij} = \mu + T_i + B_j + BT_{ij} + \varepsilon_{ij},$$

where μ is the total effect in all observations or the true average of an ensemble to which the sample belongs, T_i corresponds to the effect of the first factor at the ith level, i.e. to one of the arias, B_j is the effect of the second factor at the jth level, i.e. from the dress, BT_{ij} is the interaction effect (the singer may well feel uncomfortable in a folk costume while singing an academic aria), y_{ij} is the value of the parameter measured, and

lastly, ε_{ij} is the random error of the experiment. But the meaning of the model is different here from that of the regression equation, since it is just a formula for calculating the theoretical values of the parameter being measured at individual (discrete) points, the points of our design.

It is worth noting that when factors are independent, the variance of the parameter to be measured equals the sum of variances of the terms. Using this remarkable feature of variance, we can go on with our analysis and examine the contributions of each factor, to estimate the relative importance of each of them, to optimize their combination. This theory is called the *analysis of variance.*

We will consider the application of this analysis again referring to our singer.

All in all, we have $5 \times 3 \times 5 = 75$ combinations possible. The combinations are best represented in tabular form. In a table for each audience we will have arias arranged in a column and dresses in lines (Table 10).

Table 10

Complete Factorial Design for an Audience

	A	*B*	*C*
1	*	*	*
2	*	*	*
3	*	*	*
4	*	*	*
5	*	*	*

Instead of the asterisks the results obtained must be entered, i.e. the values of y for an aria and a dress, corresponding to the line and column of the place.

But the singer will hardly be able to make 75 performances when there is only a month to go to the contest, she can at best put in one fifth of the figure. And here again her mathematician friend comes up with a suggestion: each aria should be sung at least once in each of the dresses, and each dress must be shown off at least once in each of the audiences. Now only arias and audiences remain unbalanced, and so the strategy is termed the partially balanced design (Table 11).

Table 11

Partially Balanced Incomplete Block Design

	A	*B*	*C*
1	α	β	γ
2	β	γ	δ
3	γ	δ	ε
4	δ	ε	α
5	ε	α	β

Each entry here recommends to perform an experiment to realize the combination of a dress (column), aria (line) and audience (Greek letter). The design thus contains 15 experiments.

After an experiment the result is entered into the block and then the entire table is processed.

The analysis of variance is quite an effective technique of comparing similar combinations to

choose the best one. So a similar design has been used in the lubricant additive problem, but already not to determine optimal concentrations but to select the best composition of additives in a situation with five antioxidation additives, three antiwear additives, and five anticorrosive additives. Without statistics we would have to test all 75 combinations possible.

Statistical methods illustrated in Table 11 not only allowed the number of experiments to be reduced to 15 (without any loss of completeness), but also allowed to reveal much important evidence concerning the behaviour of the additives and their interaction, including such factors which would be impossible to establish without statistics.

Analysis of variance is widely used in psychology, biology, chemistry—virtually everywhere where qualitative factors are involved.

A Helper—the Chance

Staking money on a horse, crossing a street, and plunging into a marriage after a two month's acquaintance you mostly rely on a run of luck. Also, there are situations where without the intervention of chance you either will have no event at all, or it is precisely luck that makes it possible to tackle the problems at hand.

So chance is the only mechanism responsible for adequate functioning of ultrashort wave radio links. Long and medium waves mostly used by broadcast stations may go round the curvature of earth, short waves reflect from the ionosphere,

but ultrashort waves penetrate the ionosphere and do not follow the terrestrial curvature. So ultrashort waves propagate virtually along the line of sight, just like rays of light. At the same time, the ultrashort wave range has some attractions for a number of physical and technical reasons.

But despite the established line-of-sight propagation of ultrashort waves some anomalies have been witnessed. For example programmes of the Belgian television were once received in the USSR.

The phenomenon can be explained as follows. The lower atmosphere, called the troposphere, is continually in the state of turbulent motion, i.e. disorderly eddies of air occur. Some eddies can be seen when observing chimney smoke: smoke normally ascends following meandering fancy trajectories. The turbulence comes from a wide variety of causes, such as winds, air currents, inhomogeneity of heating of various areas of earth by sun, and so on.

The turbulent motion is responsible for random variations in air density, temperature and humidity, which in turn produce fluctuations of the refractive index and dielectric constant of the air.

This turbulence of the troposphere is modelled by an ensemble of scattering centres. A scattering centre may be pictured as a ball, and a multitude of such balls randomly arranged in space, represent our model of the tropospheric inhomogeneities. When a wave from the transmitter is incident on a region of turbulent troposphere, a flux of omnidirectionally scattered energy emerges. To be sure, the bulk of the energy travels on along

the initial direction of the wave, but some of the energy is reflected and comes to the input of the receiver, whose aerial may be positioned in the shadow of the transmitter. Figure 33 illustrates the situation.

Radio links are now available that make use of the long-distance tropospheric scattering of ultrashort wave, and the random mechanism of wave

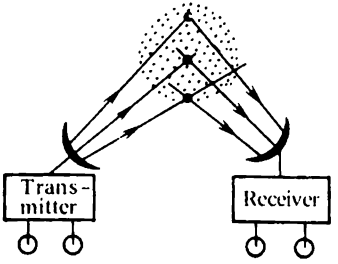


Fig. 33

scattering in the turbulent troposphere is the only mechanism on which the link relies for its operation. Should the troposphere "quiet down" and turbulent fluctuations discontinue, then the ultrashort wave radio link would stop its functioning, since no signals would come to the receiving aerial. So a random mechanism lies at the very foundation of signal transmission.

Modelling of processes relating the response function with the factors is complicated by unknown, and at times even known variables, such as the states of the object that are hard to control.

But difficulty is not impossibility. But how can we get rid of those variables? Of course, by separating the dependence under study from in-

terferences, we should make this with permissible accuracy and if such a separation be impossible to be carried out accurately.

In most of tangled situations it is, however, impossible to separate "useful" and "interfering" variables. But even when the states of the object and inputs are independent, and when they can be separated and studied one after another, it is virtually impossible to take into account all the combinations. For example, in the primary refining of oil, it is separated into several fractions: petrols, jet and diesel fuels, gas oil, lubricants, tar, etc.—more than a dozen fractions all in all. Even if each of the fractions was defined by two numbers only, the upper and lower levels, we would have about 2^{10} combinations. But in actual practice, each of the components is described by many numbers, and so the number of combinations grows beyond belief. But the main things here are the working conditions of the object described by temperatures and pressures at various locations throughout an enormous installation, raw material flow rates, and so on and so forth. Even if these quantities had only two levels, we would end up with about 2^{100}, or more than 10^{30}, combinations. No computers, whatever their speed, can handle these astronomic numbers.

Consequently, any attempts to get rid of spurious variables by meticulously cataloguing them are again "the devil's plot", since this is a way to the unifactorial experiment with all the values of all the variables fixed.

This, of course, does not make the life of the experimentalist easier. But there is a way out

here: instead of cataloguing all the variations, we should take advantage of chance, put it to our service.

Suppose we sit in at a session of a psychologist who times the solving of arithmetic problems of some type by schoolchildren. The test group includes five boys and five girls.

Previous experiments do not warrant a conclusion as to who are better at sums, boys or girls. In what order must we subject the children to test?

We may begin by testing girls—ladies first. But the children have just come from sports, and girls are more tired than boys. In addition, some of the children are older than ten years and the others are younger, their academic marks are different too, and so on. In what order then are they to be tested?

The answer to the situation is to provide random character of the test. Randomization helps to average out the effects due to fatigue, spread in their ages and IQ's.

Similar problems arise all the time in experimental biology and medicine, when some toxic preparations, radiations, or cures are tested on a group of white mice or guinea pigs. Biologists normally divide the animals into two groups: tested and controls. Say, we have 40 white mice. How are they to be divided into two groups, how many are to be taken into each group, in what order are they to be subjected to the doses? At first glance, the tiny animals appear similar, and so it would seem that any way of dividing them would do. For example, to divide the lot into two and then take one after another.

But it was repeatedly found that in a situation when it seems to a man that his choice is arbitrary, his actions are nevertheless purposeful. Involved here are some vague mechanisms of subconscious activity. A biologist sometimes subconsciously selects for his experiments a group of weaker animals, when he seeks to prove the effectiveness of a poison, or on the contrary stronger animals, when he wants to prove the effectiveness of his antidote. It should be stressed that the experimentalist should not be blamed with conscious manipulations, by no means! Only subconscious mechanisms are involved here.

My friend once discussed with me an amazing fact, we tried then to account for it together. He got computer assistants to work out firing tables for one of the arms he worked on at the time. The resultant table contained ten thousand five-digit numbers. He looked over the pages, took at random a number, checked the calculations and found a mistake. He then took another number, again at random, and again found a mistake. He was irate and made the assistants to recalculate the entire table, checking and rechecking. All the remaining 9,998 numbers turned out to be correct!

He was dead sure that his selection was "random", and so he considered the selection of any of these 10,000 numbers equiprobable. The probability of selecting just these two wrong numbers is 2/10,000, an exceedingly small number, and so what happened was unlikely.

Now, when we look back at the amazing happening, one thing is clear. Being a very experienced specialist, my friend singled out numbers

that deviated, if only slightly, from what could be expected. But the process was entirely subconscious, he could not explain what prompted him to take these numbers.

Consequently, his choice was by no means random, and by no means equiprobable event. How then can we provide a really random selection of children, numbers of guinea pigs, such that is free of arbitrary judgement of the experimenter. For this purpose, tables of random numbers are used. These are compiled readily. Get ten identical balls, mark them with the digits from zero to nine and place them in a bag. After careful strirring, take out one, write the number and return it into the bag. Again stir the balls and go through the procedure once more—you thus get a second number. Reiterating the procedure you will obtain a table of random numbers. So you can get columns of numbers with any number of gidits.

Of course, in real life nobody uses this procedure, but still many real techniques that rely on fast computers, in essence, model this procedure.

When a random order is required for an experiment, objects must at first be assigned numbers (e.g. guinea pigs), and then split into groups using tables of random numbers. Suppose you assign numbers to each of your 40 guinea pigs. Then select into your control group the first 20, whose two-digit numbers turn up among the first pairs of the table (numbers larger than 40 are rejected). In the same way, the experimental group is gathered.

Likewise, the sequence of experiments in the lubricant additive problem should be random.

And the performances of the singer: the entries in the partially balanced incomplete block design (Table 11) should be made using a random numbers table.

To summarize, the random order of selecting the values of the factors is a reliable way of avoiding prejudicing the experiment. And today's statistician preparing an experiment randomizes his experiment, relying on chance as his ally.

Concluding Remarks

Let us leaf the book. What is it about? About statistical hypothesis testing and sequential analysis, about the theory of risk and modelling, about identification and prediction, about the theory of passive and active experiment. This parade of technical terms might suggest that the book is devoted to esoteric issues. But my intentions were different. Whatever your field, reader, in your work and everyday life you appear as experimentalist, observer and decision-maker. Sometimes you tackle your problems, make observations or experiments in a difficult situation, in a fog of uncertainty. And it is here that statistical methods may come in handy.

Experimental design enables the experimentalist to save efforts and funds. But this is not all there is to it. Conscious design necessitates clear understanding of the entire procedure that involves the conception of the experiment, analysis of a priori information, modelling, before randomization, sequential strategy, optimal selection of point in a factor space, reasonable inter-

pretation of results of statistical processing, represented in a compact, convenient form.

The art of research is the art of modelling. And if the situation calls for randomization, the investigator must be able to employ an effective statistical method. To this end, one must not only have some grounding in statistics, but must also feel the problem and snags involved. Both arts do not come easily, therefore a skilled statistician on a team of experimentalists will be of help.

To the Reader

Mir Publishers would be grateful for your comments on the content, translation and design of this book.

We would also be pleased to receive any other suggestions you may wish to make.

Our address is:

Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

Printed in the Union of Soviet Socialist Republics

Also from Mir Publishers

L. A. Rastrigin Contemporary Principles to Control Complex Objects

This book is the first in the world to attempt a "thorough" presentation of the contemporary principles, approaches, and methods to control complex, but formal, objects of diverse physical natures.

The book is an intermediate one between a scientific monograph and popular science. It introduces the reader to the problem of controlling complex objects and is in a form accessible to an engineer.

A. I. Kitaigorodsky Physics for Everyone

This is the fourth and concluding book in the *Physics for Everyone* series by the world-renowned scientist, winner of the Nobel and Lenin prizes, academician Lev Landau, and the distinguished physicist Alexander Kitaigorodsky. This book discusses in a simple and easy-to-understand manner the phenomena of electromagnetic waves, thermal radiation, and current treatment of spectroscopic analysis. Provides an introduction to the field of nuclear physics and explains the most common types of lasers. Outlines principal aspects of special theory of relativity and quantum mechanics.

M. I. Kaganov and I. M. Lifshits Quasiparticles

This book, *Quasiparticles* (Ideas and principles of solid state quantum physics), was written by two outstanding specialists in the solid state quantum theory. The book is meant for a broad circle of readers. It contains no complicated mathematical formulas, nor derivations of them. The authors base the presentation on analogies, assuming their readers to possess certain degree of intuition in the field of physics. The book tells how atomic particles move inside solids, what is meant by the term "thermal motion", and how the characteristics of motion of atomic particles are revealed in the macroscopic properties of solids.

B. N. Pshenichny and **Yu. M. Danilin** Numerical Methods in Extremal Problems

The book describes methods and algorithms for numerical solution of problems of finding extrema of functions and functionals met with in mathematical programming. economics, optimal control theory and other fields of science and practice.

Special attention is paid to algorithms with a fast rate of convergence and implementable on computers. Methods of unconstrained and constrained minimization of functions of independent variables are discussed. The book will be useful to specialists in mathematical programming, computational mathematics and optimal control theory and to broad circles of students and engineers, who in their practical work have to solve problems of function minimization.

SCIENCE FOR EVERYONE

Professor Yakov Khurgin is known for his works in applied mathematics, cybernetics, industrial applications of mathematics. Besides he is an imaginative popularizer of science. His book Do You Say Mathematics? has been passed through two editions in Russian and has been translated into English. His new book Yes, No or Maybe is a readable introduction to the frontiers of statistics and experiment design. The book is written in a lovely language and contains many practical examples.