XIV

CHANCE, THE EVER-PRESENT RIVAL CONJECTURE

. . . the probability that this coincidence is a mere work of chance is, therefore, considerably less than $(1/2)^{60}$ *. . . Hence this coincidence must be produced by some cause, and a cause can be assigned which affords a perfect explanation of the observed facts.*—G. KIRCHHOFF[1]

1. Random mass phenomena. Everyday speech uses the words "probable," "likely," "plausible," and "credible" in meanings which are not sharply distinguished. Now, we single out the word "probable" and we shall learn to use this word in a specific meaning, as a technical term of a branch of science which is called the "Theory of Probability."[2]

This theory has a great variety of applications and aspects and, therefore, it can be conceived and introduced in various ways. Some authors regard it as a purely mathematical theory, others as a kind or branch of logic, and still others as a part of the study of nature. These various points of view may or may not be incompatible. We have to start by studying one of them, but we should not commit ourselves to any of them. We shall change our position somewhat in the next chapter, but in the present chapter we choose the viewpoint which is the most convenient in the great majority of applications and which the beginner can master most quickly. We regard here the theory of probability as a part of the study of nature, as the theory of certain observable phenomena, the *random mass phenomena.*[3] We can understand pretty clearly what this term means if we compare a few familiar examples of such phenomena.

[1] *Abhandlungen der k. Akademie der Wissenschaften*, Berlin, 1861, p. 79.

[2] In the foregoing, the words "probable" and "probability" have been sometimes used in a non-technical sense, but this will be carefully avoided in the present chapter and the next. The words "likely" and "likelihood" will be introduced as technical terms later in this chapter.

[3] In this essential point, and in several other points, the present exposition follows the views of Richard von Mises although it deviates from his definition of mathematical probability; cf. his book *Probability, Statistics, and Truth.*

(1) *Rainfall.* Rainfall is a mass phenomenon. It consists of a very great number of single events, of the fall of a very great number of raindrops. These raindrops, although very similar to each other, differ in various respects: in size, in the place where they strike the ground, etc. There is something in the behavior of the raindrops that we properly describe as "random." In order to understand clearly the meaning of this term let us imagine an experiment.

Let us observe the first drops on the pavement as the rain starts falling. We observe the pavement in the middle of some large public square, sufficiently far from buildings or trees or anything that could obstruct the rain. We focus our attention on two stones which we call the "right-hand stone" and the "left-hand stone." We observe the drops falling on these stones and we note the order in which they strike. The first drop falls on the left-hand stone, the second drop on the right, the third again right, the fourth left, and so on, without apparent regularity as

$$L\ R\ R\ L\ L\ L\ R\ L\ R\ L\ R\ R\ L\ R\ R$$

(R for right, L for left). There is no regularity in this succession of the raindrops. In fact, having observed a certain number of drops, we cannot reasonably predict which way the next drop will fall. We have noted above fifteen entries. Looking at them, can we predict what the sixteenth entry will be, R or L? Obviously, we cannot. On the other hand, there is some sort of regularity in the succession of the raindrops. In fact, we can confidently predict that at the end of the rain the two stones will be equally wet. That is, the number of drops striking each stone will be very nearly proportional to the area of its free horizontal surface. Nobody doubts that this is so, and the meteorologists certainly assume that this is so in constructing their rain-gauges. Yet there is something paradoxical. We can foresee what will happen in the long run, but we cannot foresee the details. The rainfall is a typical random mass phenomenon, *unpredictable in certain details, predictable in certain numerical proportions of the whole.*

(2) *Boys among the newborn.* In a hospital, the newborn babies are registered in order as they are born. Boys and girls (B and G) follow each other without apparent regularity as

$$G\ B\ B\ G\ B\ G\ B\ B\ G\ G\ B\ B\ B\ G\ G$$

Although we cannot predict the details of this random succession, we can well predict an important feature of the final result obtained by summing up all such registrations in the United States during a year: the number of the boys will be greater than the number of the girls and, in fact, the ratio of these two numbers will be little different from the ratio 51.5 : 48.5. The number of births in the United States is about 3 millions per year. We have here a random mass phenomenon of considerable dimensions.

(3) *A game of chance.* We toss a penny repeatedly, noting each time which side it shows, "heads" or "tails" (H or T). We obtain so a succession without apparent regularity as

$$T\ H\ H\ H\ T\ H\ T\ H\ H\ T\ H\ T\ H\ T\ T$$

If we have the patience to toss the penny a few hundred times, a definite ratio of heads to tails emerges, which does not change much if we prolong our experiment still further. If our penny is "unbiased," the ratio 50 : 50 of heads to tails should appear in the long run. If the penny is biased, some other ratio will come into view. At any rate we see again the characteristic features of a random mass phenomenon. Constant proportions emerge in the long run, although the details are unpredictable. There is a certain aggregate regularity, in spite of the irregularity of the individual happenings.

2. The concept of probability. In the year 1943 the number of births in the United States, male, female, and total, was

1,506,959   1,427,901   2,934,860,

respectively. We call

1,506,959 the frequency of the male births,

1,427,901 the frequency of the female births.

We call

$$\frac{1,506,959}{2,934,860} = 0.5135$$

the relative frequency of the male births and

$$\frac{1,427,901}{2,934,860} = 0.4865$$

the relative frequency of the female births. In general, if an event of a certain kind occurs in m cases out of n, we call m the *frequency* of occurrence of that kind of event and m/n its *relative frequency.*

Let us imagine that, throughout the whole year, the births are successively registered in the whole United States (as in the hospital that we have mentioned in the foregoing section). If we look at the succession of male and female births, we have before us an extremely long series of almost three million entries beginning like

$$G\ B\ B\ G\ B\ G\ B\ B\ G\ G\ B\ B\ B\ G\ G.$$

As the mass phenomenon unfolds, we have, at each stage of the observation, a certain frequency of male births and also a certain relative frequency. Let us note, after 1, 2, 3, . . . observations, the frequencies and the relative frequencies found up to that point:

Observations	*Event*	*Frequency of B*	*Relative frequency*
1	G	0	$0/1 = 0.000$
2	B	1	$1/2 = 0.500$
3	B	2	$2/3 = 0.667$
4	G	2	$2/4 = 0.500$
5	B	3	$3/5 = 0.600$
6	G	3	$3/6 = 0.500$
7	B	4	$4/7 = 0.571$
8	B	5	$5/8 = 0.625$
9	G	5	$5/9 = 0.556$
10	G	5	$5/10 = 0.500$
11	B	6	$6/11 = 0.545$
12	B	7	$7/12 = 0.583$
13	B	8	$8/13 = 0.615$
14	G	8	$8/14 = 0.571$
15	G	8	$8/15 = 0.533$

As far as we have tabulated it, the relative frequency oscillates pretty strongly (between the limits 0.000 and 0.667). Yet we have here only a very small number of observations. As we go further and further, the oscillations of the relative frequency will become less and less violent, and we can confidently expect that in the end it will oscillate very little about its final value 0.5135. *As the number of observations increases, the relative frequency appears to settle down to a stable final value, in spite of all the unpredictable irregularities of detail.* Such behavior, the emergence of a stable relative frequency in the long run, is typical of random mass phenomena.

An important aim of any theory of such phenomena must be to predict the final stable relative frequency or *long range relative frequency.* We have to consider the *theoretical value of long range relative frequency* and we shall call this theoretical value *probability.*

We wish to clarify this concept of probability. Naturally, we begin with the study of mass phenomena for which we can predict the long range relative frequency with some degree of reasonable confidence.

(1) *Balls in a bag.* A bag contains p balls of various colors among which there are exactly f white balls. We use this simple apparatus to produce a random mass phenomenon. We draw a ball, we look at its color and we write W if the ball is white, but we write D if it is of a different color. We put back the ball just drawn into the bag, we shuffle the balls in the bag, then

we draw again one and note the color of this second ball, W or D. In proceeding so, we obtain a random sequence similar to those considered in sect. 1:

$$W\ D\ D\ D\ W\ D\ D\ W\ W\ D\ D\ D\ W\ W\ D.$$

What is the long range relative frequency of the white balls?

Let us discuss the circumstances in which we can predict the desired frequency with reasonable confidence. Let us assume that the balls are homogeneous and exactly spherical, made of the same material and having the same radius. Their surfaces are equally smooth, and their different coloration influences only negligibly their mechanical behavior, if it has any influence at all. The person who draws the balls is blindfolded or prevented in some other manner from seeing the balls. The position of the balls in the bag varies from one drawing to the other, is unpredictable, beyond our control. Yet the permanent circumstances are well under control: the balls are all the same shape, size, and weight; they are *undistinguishable* by the person who draws them.

Under such circumstances we see no reason why one ball should be preferred to another and we naturally expect that, in the long run, each ball will be drawn approximately *equally often*. Let us say that we have the patience to make 10,000 drawings. Then we should expect that each of the p balls will appear about

$$\frac{10{,}000}{p} \text{ times.}$$

There are f white balls. Therefore, in 10,000 drawings, we expect to get white

$$f\,\frac{10{,}000}{p} = 10{,}000\,\frac{f}{p} \text{ times;}$$

this is the expected frequency of the white balls. To obtain the relative frequency, we have to divide the frequency by the number of observations, or drawings, that is, by 10,000. And so we are led to the statement: the long range relative frequency, or *probability*, of the white balls is f/p.

The letters f and p are chosen to conform to the traditional mode of expression. As we have to draw one of the p balls, we have to choose one of p *possible* cases. We have good reasons (equal condition of the p balls) not to prefer any of these p possible cases to any other. If we wish that a white ball should be drawn (for example, if we are betting on white), the f white balls appear to us as *favorable* cases. Hence we can describe the probability f/p as the *ratio of the number of favorable cases to the number of possible cases.*

Pulling a ball from a bag, putting the ball back into the bag, shaking the bag, pulling another ball, and repeating this n times seems to be a pretty silly occupation. Do we waste our time in studying such a primitive game?

I do not think so. The bag and the balls, handled in the described manner, generate a random mass phenomenon which is particularly simple and accessible. Generalization naturally starts from the simplest, the most transparent particular case. The science of dynamics was born when Galileo began studying the fall of heavy bodies. The science of probability was born when Fermat and Pascal began studying games of chance which depend on casting a die, or drawing a card from a pack, or drawing a ball from a bag. The fundamental concepts and laws of dynamics can be extracted from the simple phenomenon of falling bodies. We use the bag and the balls to understand the fundamental concept of probability.

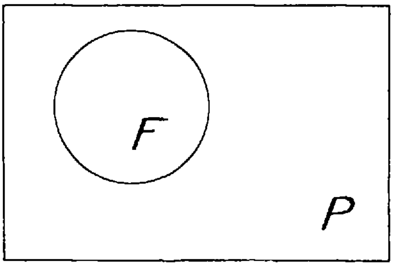


Fig. 14.1. Probability defined by rainfall.

(2) *Rainfall.* We return to the consideration of the random mass phenomenon from which we started in sect. 1. The area of a horizontal surface is P and the area of a certain portion of this surface is F; see fig. 14.1. We observe the raindrops falling on this area P and we are interested in the frequency of the raindrops falling on the subarea F. We are inclined to predict without hesitation the long range relative frequency: the fraction of the total rain over the area that falls on the subarea will be very nearly F/P if the rain consists of more than a few drops. In other words, the probability that a raindrop striking the surface of area P should strike the portion of area F is F/P. If we idealize the rainfall and consider a raindrop as a geometric point, we can also say: the probability that a point falling in the area P should fall in the subarea F is F/P.

In the last statement we consider each point of the area P as a possible case and each point of the subarea F as a favorable case. The number of favorable cases as that of the possible cases is infinite, and it would not make sense to talk about the ratio of infinite numbers. We can consider, however, the area of a surface as the *measure* of the points contained in the surface. Using this term, we can describe the probability F/P as the *ratio of the measure of favorable cases to the measure of possible cases.*

3. Using the bag and the balls. In deriving the fundamental principle of statics, Lagrange replaced an arbitrary system of forces by a suitable system of pulleys. In the light of this Lagrangeian argument (the details of which are not needed here[4]) any case of equilibrium appears as a suitable combination of correctly balanced pulleys. The Calculus of Probability can be viewed in a similar manner; in fact, such a view is suggested by the early history of this science. Seen from this standpoint, any problem of probability appears comparable to a suitable problem about bags containing balls, and any random mass phenomenon appears as similar in certain essential respects to successive drawings of balls from a system of suitably combined bags. Let us illustrate this by a few simple examples.

(1) Instead of tossing a fair penny, we can draw a ball from a bag containing just two balls, one of which is marked with an *H* and the other with a *T* (heads and tails). Instead of casting an unbiased die, we can draw a ball from a bag containing exactly six balls, marked with 1, 2, 3, 4, 5, or 6 spots, respectively. Instead of drawing a card from a pack of cards, we can draw a ball from a bag containing 52 balls, suitably marked. It seems to be intuitively clear that substituting a bag with balls for pennies, dice, cards, and other similar contrivances in a suitable way, we do not change the odds in the usual games of chance. At least, we do not change the chances in that idealized version of these games in which the contrivances used (pennies, dice, etc.) are supposed to be perfectly symmetrical and, correspondingly, certain fundamental chances perfectly equal.

(2) Wishing to study the randomness in the distribution of boys and girls among the newborn, we may substitute for the actual mass phenomenon successive drawings from a bag containing 1,000 balls, 515 marked with *B* and 485 marked with *G*. This substitution is, of course, theoretical and, as every theory is bound to be, it is tentative and approximative. Yet the point is that the bag and the balls enable us to formulate a theory.

(3) A meteorologist registers the succession of rainy and rainless days in a certain locality. His observations seem to show that, on the whole, each day tends to resemble the foregoing day: rainless days seem to follow rainless days more easily than rainy days and, similarly, rainy days seem to follow rainy days more easily than rainless days. Of course, a dependable regularity appears only in a long series of observations; the details are irregular, seem to be random.

The meteorologist may wish to express more clearly his impressions that we have just sketched. If he wishes to formulate a theory in terms of probability, he may consider three bags. Each bag contains the same number of balls, let us say 1,000 balls. Some of the balls are white, the others are black (white for rainless, black for rainy). Yet there are important differences between the bags. Each bag bears an inscription, easily visible to the person who draws the balls. One bag is inscribed "START," another "AFTER WHITE," and the third "AFTER BLACK." The ratio of balls of different color is different in different bags. In each bag the ratio of white balls to

[4] See E. Mach, *Die Mechanik*, p. 59–62.

black balls approximates the observable ratio of rainless days to rainy days, but in different circumstances. In the bag "START" the ratio is that of rainless days to rainy days throughout the year, in the bag "AFTER WHITE" the ratio is that of rainless days to rainy days following a rainless day, and in the bag "AFTER BLACK" the ratio is that of rainless days to rainy days following a rainy day. Therefore, the bag "AFTER WHITE" contains *more* white balls than the bag "AFTER BLACK." The balls are drawn successively and each ball drawn, when its color has been noticed, is replaced into the bag from which it was drawn. The bag "START" is used but once, for the first ball. If the first ball is white, we use the bag "AFTER WHITE" for the second ball, but if the first ball is black, the second ball is drawn from the bag "AFTER BLACK." And so on, the color of the ball just drawn determines the bag from which the next ball should be drawn.

It is just a theory that the succession of white and black balls drawn under the described circumstances imitates the succession of rainless and rainy days with a reasonable approximation. Yet, on the face of it, this theory does not seem to be out of place. At any rate, this theory, or some similar theory, could deserve to be confronted with the observations.

(4) Take any English text (from Shakespeare, if you prefer) and replace each of the letters *a*, *e*, *i*, *o*, *u*, and *y* by *V* and each of the remaining twenty letters by *C*. (*V* means vowel and *C* means consonant.) You obtain a pattern as

$$C\ V\ C\ V\ V\ C\ C\ V\ C\ C\ V\ C\ V\ C\ C\ .$$

This irregular sequence is in some way opposite to that discussed in the foregoing subsection (3): each day tends to be like the foregoing day, but each letter tends to be unlike the foregoing letter. Still, we could imitate the succession of vowels and consonants by a succession of white and black balls drawn from three bags bearing the same inscriptions as before (in subsection (3)), yet the ratio of white balls to black balls should not be the same as before. To imitate realistically the succession of vowels and consonants the bag "AFTER WHITE" should contain *less* white balls than the bag "AFTER BLACK."

(5) There are two bags. The first bag contains p balls among which there are f white balls. The second bag contains P chips among which there are F white chips. Using both hands, I draw from both bags at the same time, a ball with the left hand and a chip with the right hand. What is the probability that both the ball and the chip turn out to be white?

We could, of course, repeat this primitive experiment sufficiently often, perhaps a thousand times, and so obtain an approximate value for the desired probability. Yet we can also try to guess it, and that is more interesting.

The result of the two simultaneous drawings is a "couple," consisting of a ball and a chip. There are p balls and P chips. As any ball can be coupled with any chip, there are pP possible couples; they are shown in fig. 14.2 where $p = 5, f = 2, P = 4, F = 3$. There is no reason to prefer any of the p balls to any other ball, or any of the P chips to any other chip. There seems to be *no reason to prefer any of the pP couples to any other couple.* In fact, in performing the experiment with the two bags, I am supposed to proceed blindly, at random, so that each hand draws independently of the other. "Let not thy left hand know what thy right hand doeth." It seems incredible that the chances of the ball that I draw with my left hand should be influenced by the chip that I draw with my right hand. Why should ball no. 1 be any more attracted by chip no. 1 than it is by chip no. 2?

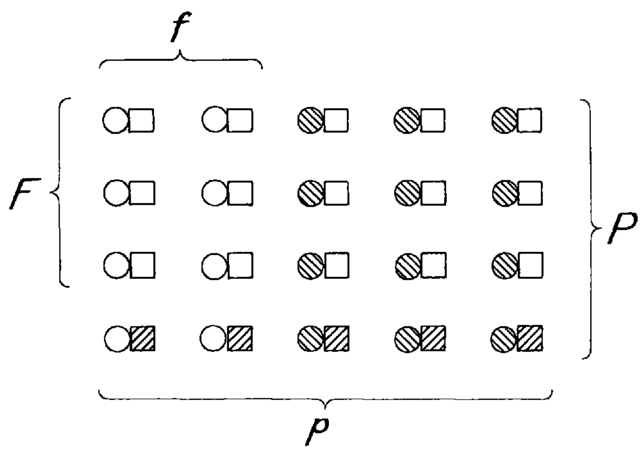


Fig. 14.2. Independent events.

And so we can imagine a bag, containing pP mechanically indistinguishable objects (each object is a couple, a ball attached to a chip); one drawing from this one bag appears *equivalent* to the two simultaneous drawings from the two bags described at the outset. We have so pP possible cases; it remains to find the number of favorable cases. A glance at fig. 14.2 shows that there are fF couples consisting of a white ball and a white chip. And so we obtain the value of the desired probability: it is

$$\frac{fF}{pP} = \frac{f}{p} \cdot \frac{F}{P}$$

the *product* of two probabilities. In fact, f/p is the probability of drawing a white ball from the first bag, and F/P the probability of drawing a white chip from the second bag.

The essential point about the ball and the chip is that the drawing of one does not influence the chances of the other. In the usual terminology of the calculus of probability, such events are called *independent* of each other; the joint happening of both events is viewed as a *compound* event. The foregoing consideration motivates the rule: *The probability of a compound event is the product of the probabilities of the constituent events, provided that these constituent events are mutually independent.*

4. The calculus of probability. Statistical hypotheses. The theory of probability, as we see it, is a part of the study of nature, the theory of random mass phenomena.

The most striking achievement of the physical sciences is prediction. The astronomers predict with precision the eclipses of the sun and the moon, the position of the planets, and the return of comets which evade observation for several years. A great astronomer (Leverrier) succeeded even in predicting the position of a planet (Neptune), the very existence of which was not known before. The theory of probability predicts the frequencies in certain mass phenomena with some amount of success.

The astronomers base their predictions on former observations, on the laws of mechanics, the law of gravitation, and long difficult computations. Any branch of physical science bases its predictions on some theory or, we can say, on some conjecture, since no theory is certain and so every theory is a more or less reasonable, more or less well-supported, conjecture. In trying to predict the frequencies in a certain random mass phenomenon from the theory of probability we have to make some theoretical assumption about the phenomenon. Such an assumption, which has to be expressed in terms of probability concepts, is called a *statistical hypothesis.*

When we apply the theory of probability we have to compute probabilities (which are theoretical, approximate values of relative frequencies). When we try to find a probability, we have a problem to solve. The unknown of this problem is the desired probability. Yet, in order to determine this unknown, we need data and conditions in our problem. The data are usually probabilities and the conditions, on which the relation of the unknown probability to the given probabilities depends, constitute a statistical hypothesis.

As in the applications of the theory of probability the computation of probabilities plays a prominent role, this theory is usually called the *calculus of probability.* Thus, the aim of the calculus of probability is to compute new probabilities on the basis of given probabilities and given statistical hypotheses.

The reader who wishes to peruse the remaining part of this chapter must either know the elements of the calculus of probability, or he must take for granted certain results derived from these elements. Most of the time, the text will state the results without derivation; derivations will be given subsequently, in the First Part of the Exercises and Comments following this chapter, and in the corresponding Solutions. Yet even if the reader does not check the derivation of the results, he ought to have some insight into the underlying theoretical assumptions. We can make such assumptions intuitively understandable: we compare the random mass phenomenon that we examine to drawings from suitably filled bags under suitable conditions, as in the foregoing sect. 3.

The applications of the calculus of probability are of unending variety. The following sections of this chapter attempt to illustrate the principal types of applications by suitable elementary examples. Stress will be laid on the motivation of these applications, that is, on such preliminary considerations as make the choice of procedure plausible.

5. Straightforward prediction of frequencies. At the beginning of its history the calculus of probability was essentially a theory of certain games of chance. Yet the predictions of this theory were not tested experimentally on a large scale until modern times. We begin by discussing an experiment of this kind.

(1) W. F. R. Weldon cast 12 dice 26,306 times, noting each time how many of these 12 dice have shown more than four spots.[5] The results of his observations are listed in column (4) of Table I; column (1) shows the number of the dice among the 12 that have turned up five or six spots. Thus, in 26,306 trials it never happened that all twelve dice showed more than four spots. The most frequent case was that in which four out of the twelve dice showed five or six spots; this happened 6,114 times.

Table I

(1) Nr. of 5 or 6	(2) Excess I	(3) Predicted I	(4) Observed	(5) Predicted II	(6) Excess II
0	+ 18	203	185	187	+ 2
1	+ 67	1216	1149	1146	− 3
2	+ 80	3345	3265	3215	− 50
3	+ 101	5576	5475	5465	− 10
4	+ 159	6273	6114	6269	+ 155
5	− 176	5018	5194	5115	− 79
6	− 140	2927	3067	3043	− 24
7	− 76	1255	1331	1330	− 1
8	− 11	392	403	424	+ 21
9	− 18	87	105	96	− 9
10	− 1	13	14	15	+ 1
11	− 3	1	4	1	− 3
12	0	0	0	0	0
Total	0	26,306	26,306	26,306	0

How can the theory predict the observed numbers listed in column (4) of Table I? If we assume that the dice are "fair" and that the trials with different dice, or with the same die at different times, are independent of

[5] *Philosophical Magazine*, ser. 5, vol. 50, 1900, p. 167–169; in a paper by Karl Pearson.

each other, we can compute the relevant probabilities. Under our assumption (which is properly termed a "statistical hypothesis") the probability that exactly 4 dice out of 12 should show 5 or 6 spots is

$$P = 495 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^8 = \frac{126{,}720}{531{,}441}.$$

Now, by definition, the probability is the theoretical value of long range relative frequency. If the event with probability P shows itself m times in n trials, we expect that

$$\frac{m}{n} = P \text{ approximately}$$

or

$$m = Pn \text{ approximately.}$$

Therefore, we should expect that exactly 4 dice will show five or six spots out of the 12 dice cast in about

$$Pn = \frac{126{,}720}{531{,}441} 26{,}306 = 6{,}273$$

cases out of $n = 26{,}306$ trials. (Observe that we can compute this number 6,273 before the trials start.) Now, this predicted value 6,273 does not seem to be "very different" from the observed number 6,114, and so our first impression about the practical applicability of the theory of probability may be quite good.

The number 6,273 is listed in column (3) of Table I at the proper place, in the same row as the number 4 in column (1). All the numbers in column (3) are similarly computed. In order to compare more conveniently the predicted values in column (3) with the observed numbers in column (4), we list the differences (predicted less observed) in column (2). With their meaning in mind, we survey the columns (2), (3), and (4). Is the agreement between experience and theory satisfactory? Are the observed numbers sufficiently close to the predicted values?

There is, obviously, some agreement between the columns (3) and (4). Both columns of numbers have the same general aspect: the maximum is attained at the same point (in the same row) and the numbers first increase to the maximum and then decrease steadily to 0 in very much the same fashion in both columns. The deviation of the observed number from the predicted value appears relatively small in most cases; the agreement, at a first glance, looks quite good. On the other hand, however, the number of trials, 26,306, appears pretty large. Are the deviations sufficiently small in view of the large number of trials?

This seems to be the right question. Yet we cannot answer it off-hand; we had better postpone it till we know a little more; see sect. 7 (3). Yet without any special knowledge, just with a little common sense, we can draw quite a sharp conclusion from Table I. A physicist would easily notice the following point about the columns (3) and (4). The differences are listed in column (2). Some of these differences are positive, others negative. If these differences were randomly distributed, the + and − signs should be intermingled in some disorderly fashion. In fact, however, the + and − signs are sharply separated: the theoretical values are too large up to a certain point, and too small from that point onward. In such a case, the physicist speaks of a *systematic* deviation of the theory from the experiment, and he regards such a systematic deviation as a grave objection against the theory.

And so the agreement between the theory of probability and Weldon's observations, which looked quite good at first, begins to look much less good.

(2) Yet who is responsible for that systematic deviation? The theoretical values have been computed according to the rules of the calculus of probability on the basis of a certain assumption, a "statistical hypothesis." We need not blame the rules of the calculus; the fault may be with the statistical hypothesis. In fact, this statistical hypothesis has a weak point: we assumed that the dice used in the experiment were "fair." When gentlemen play a game of dice, they should assume that the dice are fair, but for a naturalist such an assumption is unwarranted.

In fact, let us look at the example of the physicist. Galileo discovered the law of falling bodies that we write today in the usual notation as an equation:

$$s = gt^2/2;$$

s stands for space (distance), t for time. More exactly, Galileo discovered the *form* of the dependence of s on t: the distance is proportional to the square of the time t. Yet he made no theoretical prediction about the constant g that enters into this proportionality; the suitable value of g has to be found by experiments. In this respect, as in many other respects, natural science followed the example of Galileo; in countless cases the theory yielded the general form of a natural law, and the experiment had to determine the numerical values of the constants that enter into the mathematical expression of the law. And this procedure works in our example, too.

If a die is "fair," none of the six faces is preferable to the others, and so the probability for casting 5 or 6 spots is

$$\frac{2}{6} = \frac{1}{3}.$$

Even if the die is not fair, there is a certain probability p for casting 5 or 6 spots; p may be different from 1/3. (Yet not very different in an ordinary die, otherwise we would consider the die as "loaded.") We take p as a constant that has to be determined by experiment. And now, we modify our original statistical hypothesis: we *assume* that all twelve dice used have the *same probability* p for showing 5 or 6 spots. (This is a simple assumption but, of course, pretty arbitrary. We cannot believe that it is exactly true; we can only hope that it is not very far from the truth. There is virtually no chance that the dice are exactly equal, but they may be only slightly different.) We keep unchanged the other part of our former statistical hypothesis (different dice and different trials are considered as independent).

On the basis of this new statistical hypothesis we can again assign theoretical values corresponding to the observations listed in column (4) of Table I. For example, the theoretical value corresponding to the observed value 6,114 is

$$495\, p^4\, (1 - p)^8\, 26{,}306;$$

it depends on p, and also the theoretical values corresponding to the other numbers in column (4) depend on p.

It remains to determine p from the experiments that we are examining. We cannot hope to determine p from experiments exactly, only in some reasonable approximation. If we change our standpoint for a moment and consider the casting of a single die as a trial,

$$12 \times 26{,}306 = 315{,}672$$

trials have been performed; this is a very large number. The frequency of the event "five or six spots" can be easily derived from the column (4) of Table I. We find as the value of the relative frequency

$$\frac{106{,}602}{315{,}672} = 0.3376986;$$

we take this relative frequency, resulting from a very large number of trials, for the value of p. (We assume so for p a value slightly higher than 1/3.)

Once p is chosen, we can compute theoretical values corresponding to the observed frequencies. These theoretical values are tabulated in column (5) of Table I. Thus the columns (3) and (5) give theoretical values corresponding to the same observed numbers, but computed under different statistical hypotheses. In fact, the two statistical hypotheses differ only in the value of p; column (3) uses $p = 1/3$, column (5) uses the slightly higher value derived from the observations. (Column (3) can be computed before the observations, but column (5) cannot.) The differences between corresponding items of columns (5) and (4) are listed in column (6).

There is little doubt that the theoretical values in column (5) fit the observations much better than those in column (3). In absolute value, the differences in column (6) are, with just one exception, less than, or equal to, the differences in column (2) (equal in just three cases, much less in most cases). In opposition to column (2), the signs + and − are intermingled in column (6), so that they yield no ground to suspect a systematic deviation of the theoretical values in column (5) from the experimental data in column (4).

(3) Judged by the foregoing example, the theory of probability seems to be quite suitable for describing mass phenomena generated by such gambling devices as dice. If it were not suitable for anything else, it would not deserve too much attention. Let us, therefore, consider one more example.

As reported by the careful official Swiss statistical service, there were exactly 300 deliveries of triplets in Switzerland in the 30 years from 1871 to 1900. (That is, 900 triplets were born. In talking of deliveries, we count the mothers, not the babies.) The number of all deliveries (some of triplets, some of twins, most of them, of course, of just one child) during the same period in the same geographical unit was 2,612,246. Thus, we have here a mass phenomenon of considerable proportions, but the event considered, the birth of triplets, is a *rare event*. The average number of deliveries per year is

$$2,612,246/30 = 87,075,$$

the average number of deliveries of triplets only

$$300/30 = 10.$$

Of course, the event happened more often in some years, in others less often than the average 10, and in some years exactly 10 times. Table II gives

Table II

Triplets born in Switzerland 1871–1900.

(1) Deliveries	(2) Years obs.	(3) Years theor.	(4) (2) cumul.	(5) (3) cumul.
0	0	0.00	0	0.00
1	0	0.00	0	0.00
2	0	0.09	0	0.09
3	1	0.21	1	0.30
4	0	0.57	1	0.87
5	1	1.14	2	2.01
6	1	1.89	3	3.90
7	5	2.70	8	6.60
8	1	3.39	9	9.99
9	4	3.75	13	13.74
10	4	3.75	17	17.49
11	4	3.42	21	20.91
12	3	2.85	24	23.76
13	2	2.16	26	25.92
14	1	1.59	27	27.51
15	2	1.02	29	28.53
16	0	0.66	29	29.19
17	1	0.39	30	29.58
18	0	0.21	30	29.79
19	0	0.12	30	29.91

the relevant details in column (2). We see there (in the row that has 10 in the first column) that there were in the period considered exactly 4 years in which exactly 10 deliveries of triplets took place. As the same column (2) shows, no year in the period had less than 3 such deliveries, none had more than 17, and each of these extreme numbers, 3 and 17, turned up in just one year.

The numbers of column (2) seem to be dispersed in some haphazard manner. It is interesting to note that the calculus of probability is able to match the irregular looking observed numbers in column (2) by theoretical numbers following a simple law; see column (3). The agreement of columns (2) and (3), judged by inspection, does not seem to be bad; the difference between the two numbers, the observed and the theoretical, is less than 1 in absolute value, except in two cases. Yet in these two cases (the rows with 7 and 8 in the first column) the difference is greater than 2 in absolute value.

There is a device that allows us to judge a little better the agreement of the two series of numbers. The column (4) of Table II contains the numbers of column (2) "cumulatively." For example, consider the row that has 7 in column (1); it has 5 in column (2) and 8 in column (4). Now

$$8 = 0 + 0 + 0 + 1 + 0 + 1 + 1 + 5;$$

that is, 8 is the sum, or the "accumulation", of all numbers in column (2) up to the number 5, inclusively, in the respective row. (In other words, 8 is the number of those years of the period in which the number of deliveries of triplets did *not exceed* 7.) Column (5) contains the numbers of column (3) "cumulatively", and so the columns (4) and (5) are analogously derived from the observed numbers in column (2) and the theoretical numbers in column (3), respectively. The agreement between columns (4) and (5) looks excellent; the difference is less than 1 in absolute value except in just one case, where it is still less than 2.

6. Explanation of phenomena. Ideas connected with the concept of probability play a rôle in the explanation of phenomena, and that is true of phenomena dealt with by any science, from physics to the social sciences. We consider two examples.

(1) Gregor Mendel (1822–1884), experimenting with the cross-breeding of plants, became the founder of a new science, genetics. Mendel was, by the way, an abbot in Moravia, and carried out his experiments in the garden of his monastery. His discovery, although very important, is very simple. To understand it we need only the description of one experiment and an intuitive notion of probability. To make things still easier, we shall not discuss one of Mendel's own experiments, but an experiment carried out by one of his followers.[6]

Of two closely related plants (different species of the same genus) one has

[6] By Correns; see W. Johannsen, *Elemente der exakten Erblichkeitslehre*, Jena 1909, p. 371.

white flowers and the other rather dark red flowers. The two plants are so closely related that they can fertilize each other. The seeds resulting from such crossing develop into hybrid plants which have an intermediate character: the hybrids have pink flowers. (In fig. 14.3 red is indicated by more, pink by less, shading.) If the hybrid plants are allowed to become self-fertilized, the resulting seeds develop into a third generation of plants in which all three kinds are represented: there are plants with white, plants with pink, and plants with red flowers. Fig. 14.3 represents schematically the relations between the three subsequent generations.

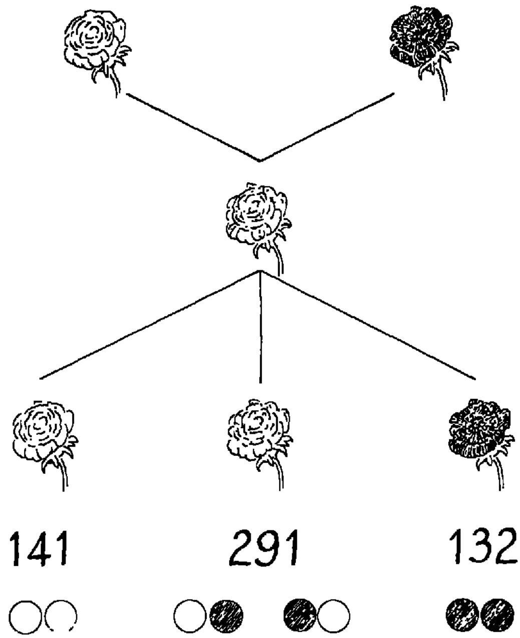


Fig. 14.3. Three generations in a Mendelian experiment.

Yet the most striking feature of the phenomenon is the numerical proportion in which the three different kinds of plants of the third generation are produced. In the experiment described, 564 plants of the third generation have been observed. Among them, those two kinds of plants that resemble one or the other grandparental plant were about equally numerous: there were 141 plants with white flowers and 132 plants with red flowers in the third generation. Yet the plants resembling the hybrid parental plants were more numerous: there were 291 plants with pink flowers in the third generation. We can conveniently survey these numbers in fig. 14.3. We easily notice that these numbers given by the experiment are approximately in a simple proportion:

141 : 291 : 132 almost as 1 : 2 : 1.

This simple proportion invites a simple explanation.

Let us begin at the beginning. The experiment began with the crossing of two different kinds of plants. Any flowering plant arises from the union of two reproductive cells (an ovule and a grain of pollen). The pink-flowering hybrids of the second generation arose from two reproductive cells of different extraction. As the pink-flowering plants of the third generation are similar to those of the second generation, it is natural to assume that they were similarly produced, by two reproductive cells of *different* kinds. This leads us to suppose that the pink-flowering hybrids of the second generation *have* two different kinds of reproductive cells. Supposing this, however, we may perceive a possibility of explaining the mixed offspring. In fact, let us see more clearly what would happen if the pink-flowering hybrids of the second generation actually *had* two different kinds of reproductive cells, which we may call "white" and "red" cells. When two such cells are combined, the combination can be white with white, or red with red, or one color with the other, and these three different combinations *could* explain the three different kinds of plants in the third generation; see fig. 14.3.

After this remark, it should not be difficult to explain the numerical proportions. The deviation of the actually observed proportion 141 : 291 : 132 from the simple proportion 1 : 2 : 1 appears as random. That is, it looks like the deviation of observed frequencies from underlying probabilities. This leads us to wondering what the probabilities of the two kinds of cells are, or in which proportion the "white" and "red" cells are produced. As there are about as many white-flowering as red-flowering plants in the third generation, we can hardly refrain from trying the simplest thing: let us assume that the "white" and "red" reproductive cells are produced in equal numbers by the pink-flowering plants. Finally, we are almost driven to compare the random encounter of two reproductive cells with the random drawing of two balls, and so we arrive at the following simple problem.

There are two bags containing white and red balls, and no balls of any other color. Each bag contains just as many white balls as red balls. With both hands, I draw from both bags, one ball from each. Find the probability for drawing two white balls, two balls of different colors, and two red balls.

As it is easily seen (cf. sect. 3 (5)), the required probabilities are

$$\frac{1}{4}, \quad \frac{2}{4}, \quad \frac{1}{4},$$

respectively. We perceive now a simple reason for the proportion 1 : 2 : 1 that seems to underlie the observed numbers, and so doing we come very close to Mendel's essential concepts.

(2) The concept of random mass phenomena plays an important rôle in physics. In order to illustrate this rôle, we consider the velocity of chemical reactions.

Relatively crude observations are sufficient to suggest that the speed of a chemical change depends on the concentration of the reacting substances. (By concentration of a substance we mean its amount in unit volume.) This dependence of the chemical reaction velocity on the concentration of the reactants was soon recognized, but the discovery of the mathematical form of the dependence came much later. An important particular case was noticed by Wilhelmy in 1850, and the general law was discovered by two Norwegian chemists, Guldberg and Waage, in 1867. We now outline, in a particular case and as simply as we can, some of the considerations that led Guldberg and Waage to their discovery.

We consider a bimolecular reaction. That is, two different substances, A and B, participate in the reaction which consists in the combination of one molecule of the first substance A with one molecule of the second substance B. The substances A and B are dissolved in water, and the chemical change takes place in this solution. The substances resulting from the reaction do not participate further in the chemical action; they are inactive in one way or another. For example, they may be insoluble in water and deposited in solid form.

The solution in which the reaction takes place consists of a very great number of molecules. According to the ideas of the physicists (the kinetic theory of matter) these molecules are in violent motion, traveling at various speeds, some at very high speed, and colliding now and then. If a molecule A collides with a molecule B, the two may get so involved that they exchange some of their atoms: the chemical reaction in which we are interested consists of such an exchange, we imagine. Perhaps it is necessary for such an exchange that the molecules should collide at a very high speed, or that they should be disposed in a favorable position with respect to each other in the moment of their collision. At any rate, the more often it happens that a molecule A collides with a molecule B, the more chance there is for the chemical combination of two such molecules, and the higher the velocity of the chemical reaction will be. And so we are led to the conjecture: *the reaction velocity is proportional to the number of collisions between molecules A and molecules B.*

We could not predict exactly the number of such collisions. We have before us a random mass phenomenon like rainfall. Remember fig. 14.2; there, too, we could not predict exactly how many raindrops would strike the subarea F. Yet we could predict that the number of raindrops striking the subarea F would be *proportional* to the number of raindrops falling on the whole area P. (The proportionality is approximate, and the factor of proportionality is F/P, as discussed toward the end of sect. 2.) Similarly, we can predict that the number of collisions in which we are interested (between any molecule A and any molecule B) will be proportional to the number of the molecules A. Of course, it will also be proportional to the number of the molecules B, and so finally proportional to the *product* of these two numbers. Yet the number of the molecules of a substance is proportional to the concentration of that substance, and so our conjecture leads us to the following statement: *the reaction velocity is proportional to the product of the concentrations.*

We arrived at a particular case of the general law of chemical mass action discovered by Guldberg and Waage. This is the particular case appropriate for the particular circumstances considered. On the basis of the law of mass action it is possible to compute the concentration of the reacting substances at any given moment and to predict the whole course of the reaction.

7. Judging statistical hypotheses. We start from an anecdote.[7]

(1) "One day in Naples the reverend Galiani saw a man from the Basilicata who, shaking three dice in a cup, wagered to throw three sixes; and, in fact, he got three sixes right away. Such luck is possible, you say. Yet the man succeeded a second time, and the bet was repeated. He put back the dice in the cup, three, four, five times, and each time he produced three sixes. 'Sangue di Bacco,' exclaimed the reverend, 'the dice are loaded!' And they were. Yet why did the reverend use profane language?"

The reverend Galiani drew a plausible conclusion of a very important type. If he discovered for himself this important type of plausible inference on the spur of the moment, his excitement is quite understandable and I, personally, would not reproach him for his mildly profane language.

The correct thing is to treat everybody as a gentleman until there is some definite evidence to the contrary. Quite similarly, the correct thing is to engage in a game of chance under the assumption that it is fairly played. I do not doubt that the reverend did the correct thing and assumed in the beginning that that man from the Basilicata had fair dice and used them fairly. Such an assumption, correctly stated in terms of probability, is a statistical hypothesis. A statistical hypothesis generally assumes the values of certain probabilities. Thus, the reverend assumed in the beginning, more or less explicitly, that any of the dice involved will show six spots with the probability 1/6. (We have here exactly the same statistical hypothesis as in sect. 5 (1).)

The calculus of probability enables us to compute desired probabilities from given probabilities, on the basis of a given statistical hypothesis. Thus, on the basis of the statistical hypothesis adopted by the reverend at the beginning, we can compute the probability for casting three sixes with three dice; it is

$$(1/6)^3 = 1/216,$$

a pretty small probability. The probability for repeating this feat twice,

[7] J. Bertrand, *Calcul des probabilités*, p. VII–VIII.

that is, casting three sixes at a first trial, and casting them again at the next trial, is

$$(1/216)^2 = (1/6)^6 = 1/46,656,$$

a very small probability indeed. Yet that man from the Basilicata kept on repeating the same extraordinary thing five times. Let us list the corresponding probabilities:

Repetitions	*Probability*		
1	$1/6^3$	=	1/216
2	$1/6^6$	=	1/46,656
3	$1/6^9$	=	1/10,077,696
4	$1/6^{12}$	=	1/2,176,782,336
5	$1/6^{15}$	=	1/470,184,984,576.

Perhaps, the reverend adopted his initial assumption out of mere politeness; looking at the man from the Basilicata, he may have had his doubts about the fairness of the dice. The reverend remained silent after the three sixes turned up twice in succession, an event that under the initial assumption should happen not much more frequently than once in fifty thousand trials. He remained silent even longer. Yet, as the events became more and more improbable, attained and perhaps surpassed that degree of improbability that people regard as miraculous, the reverend lost patience, drew his conclusion, rejected his initial polite assumption, and spoke out forcibly.

(2) The anecdote that we have just discussed is interesting in just one aspect: it is typical. It shows clearly the circumstances under which we can reasonably reject a statistical hypothesis. We draw consequences from the proposed statistical hypothesis. Of special interest are consequences concerned with some event that appears *very improbable* from the standpoint of our statistical hypothesis; I mean an event the probability of which, computed on the basis of the statistical hypothesis, is very small. Now, we appeal to experience: we observe a trial that can produce that allegedlv improbable event. If the event, in spite of its computed low probability, actually happens, it yields a strong *argument against* the proposed statistical hypothesis. In fact, we find it hard to believe that anything so extremely improbable could happen. Yet, undeniably, the thing did happen. Then we realize that any probability is computed on the basis of some statistical hypothesis and start doubting the basis for the computation of that small probability. And so there arises the argument against the underlying statistical hypothesis.

(3) As the reverend Galiani, we also felt obliged to reject the hypothesis of fair dice when we examined the extensive observations related in sect. 5 (1);

our reasons to reject it, however, were not quite as sharp as his. Could we find better reasons in the light of the foregoing discussion?

Here are the facts: 315,672 attempts to cast five or six spots with a dice produced 106,602 successes; see sect. 5 (2). *If* all dice cast were fair, the probability of a success would be 1/3. Therefore, we should expect about

$$315{,}672/3 = 105{,}224$$

successes in 315,672 trials. Thus, the observed number deviates from the expected number

$$106{,}602 - 105{,}224 = 1{,}378$$

units. Does such a deviation speak for or against the hypothesis of fair dice? Should we regard the deviation 1,378 as small or large? Is the probability of such a deviation high or low?

The last question seems to be the sensible question. Yet we still need a sensible interpretation of the short, but important, word "such." We shall reject the statistical hypothesis if the probability that we are about to compute turns out to be low. Yet the probability that the deviation should be exactly equal to 1,378 units is very small anyhow—even the probability of a deviation exactly equal to 0 would be very small. Therefore, we have to take into account *all the deviations of the same absolute value as, or of larger absolute value than, the observed deviation* 1,378. And so our judgment depends on the solution of the following problem: *Given that the probability of a success is* 1/3 *and that the trials are independent, find the probability that in* 315,672 *trials the number of successes should be either more than* 106,601 *or less than* 103,847.

With a little knowledge of the calculus of probability we find that the required probability is approximately

$$0.0000001983;$$

this means less than two chances in ten million. That is, an event has occurred that looks extremely improbable, *if* the statistical hypothesis is accepted that underlies the computation of probability. We find it hard to believe that such an improbable event actually occurred, and so the underlying hypothesis of fair dice appears extremely unlikely. Already in sect. 5 (1) we saw a good reason to reject the hypothesis of fair dice, but now we see a still better, more distinct, reason to reject it.

(4) The actual occurrence of an event to which a certain statistical hypothesis attributes a small probability is an argument against that hypothesis, and the smaller the probability, the stronger is the argument.

In order to visualize this essential point, let us consider the sequence

$$\frac{1}{10}, \quad \frac{1}{100}, \quad \frac{1}{1{,}000}, \quad \frac{1}{10{,}000}, \quad \ldots .$$

A statistical hypothesis implies that the probability of a certain event is 1/10.

The event happens. Should we reject the hypothesis? Under usual circumstances, most of us would not feel entitled to reject it; the argument against the hypothesis does not appear yet strong enough. If another event happens to which the statistical hypothesis attributes the probability 1/100, the urge to reject the hypothesis becomes stronger. If the alleged probability is 1/1,000, yet the event happens nevertheless the case against the hypothesis is still stronger. If the statistical hypothesis attributes the probability

$$\frac{1}{1{,}000{,}000{,}000}$$

to the event, or one chance in a billion, yet the event happens nevertheless, almost everybody would regard the hypothesis as hopelessly discredited, although there is no logical necessity to reject the hypothesis just at this point. If, however, the sequence proceeds without interruption so that events happen one after the other to which the statistical hypothesis attributes probabilities steadily decreasing to 0, for each reasonable person arrives sooner or later the critical moment in which he feels justified in rejecting the hypothesis, rendered untenable by its increasingly improbable consequences. And just this point is neatly suggested by the story of the reverend Galiani. The probability of the first throw of three sixes was 1/216; of the sequence of five throws of three sixes each, 1/470,184,984,576.

The foregoing discussion is of special importance for us if we adopt the standpoint that the theory of probability is a part of the study of nature. Any natural science must recur to observations. Therefore it must adopt rules that specify somehow the circumstances under which its statements are confirmed or confuted by experience. We have done just this for the theory of probability. We described certain circumstances under which we can reasonably consider a statistical hypothesis as practically refuted by the observations. On the other hand, if a statistical hypothesis survives several opportunities of refutation, we may consider it as corroborated to a certain extent.

(5) Probability, as defined in sect. 2, is the theoretical value of long range relative frequency. The foregoing gave us an opportunity to realize a few things. First, such a theoretical value depends, of course, on our theory, on our initial assumptions, on the statistical hypothesis adopted. Second, such a theoretical value may be very different from the actual value.

A suitable notation may help us to clarify our ideas. Let P be the probability of an event E computed on the basis of a certain statistical hypothesis H. Then P depends both on E and on H. (In fact, we could use, instead of P, the more explicit symbol $P(E, H)$ that emphasizes the dependence of P on E and H.)

In some of the foregoing applications we took the hypothesis H for granted (at least for the moment) and, computing P on the basis of H, we tried to predict the observable frequency of the event E. Yet, in the present section, we proceeded in another direction. Having observed the event E, we computed P on the basis of the statistical hypothesis H and, in view of the value of P obtained, we tried to judge the reliability of the hypothesis H. We perceive here a new aspect of P. The *smaller* P is, the more we feel inclined to reject the hypothesis H, and the *more unlikely* the hypothesis H appears to us: P indicates the likelihood of the hypothesis H. We shall say henceforward that P is the *likelihood of the statistical hypothesis* H, judged in view of the fact that the event E has been observed.

This terminology, which agrees essentially with the usage of statisticians, emphasizes a certain aspect of the dependence of P on the event E and the statistical hypothesis H. Our original terminology lays the stress on the complementary aspect of the same dependence: P is the *probability of the event* E, computed on the basis of the statistical hypothesis H.

Some practice in the use of this double terminology is needed to convince us that its advantages sufficiently outweigh its dangers.

8. Choosing between statistical hypotheses. The following example may provide a first orientation to the applications of the theory of probability in statistical research.

(1) A consumer buys a certain article from the producer in large lots. The consumer is a big consumer, a large merchandizing or industrial firm, or a government agency. The producer is also big and manufactures the article in question on a large scale. The article can be a nail, or a knob, or anything manufactured; an interesting example is a fuze, used for firing explosives in ammunition or in blasting operations. The article has to meet certain specifications. For example, the nail should not be longer than 2.04 inches nor shorter than 1.96 inches, its thickness is similarly specified, and perhaps also its minimum breaking strength; the burning time of the fuze is specified, and so on. An article that does not meet the specifications is considered as *defective*. Even the most carefully manufactured lot may contain a small fraction of defectives. Therefore the lot has to be inspected before it passes from the producer to the consumer. The lot may be fully inspected, that is, each article in the lot may be tested whether it meets the agreed specifications. Such a full inspection would be impractical for a lot of 10,000 nails and it would be preposterous for a lot of fuzes even if the lot is small; in order to measure its burning time, you have to destroy the fuze and there is not much point in destroying the whole lot by inspecting it. Therefore in many cases instead of inspecting the whole lot before acceptance, only a relatively small sample is taken from the lot. A simple procedure of such acceptance sampling is characterized by the following rule.

"Take a random sample of n articles from the submitted lot of N articles. Test each article in the sample. If the number of defectives in the sample does not exceed a certain agreed number c, the so-called *acceptance number*, the consumer accepts the lot, but he rejects it, and the producer takes it back, if there are more defectives than c in the sample."

The results obtained by this rule depend on chance. By chance, the fraction of defectives in the sample can be much lower or much higher than in the whole lot. If the sample turns out to be better than the lot, chance works against the consumer, and it works against the producer if the sample turns out to be worse than the lot. In spite of these risks, some such procedure appears necessary, and the rule formulated may be quite reasonable. We have to find out how the procedure works, how its result depends on the quality of the submitted lot. And so we are led to formulate the following problem: *Given p, the probability that an article chosen at random in the submitted lot is defective, find a, the probability that the lot will be accepted.*

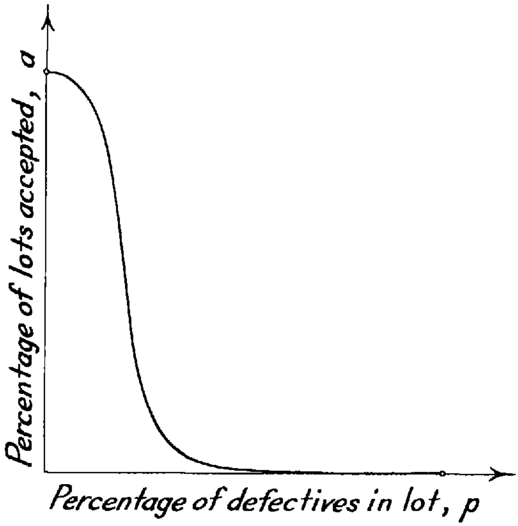


Fig. 14.4. Operating characteristic of an acceptance sampling procedure.

In the most important practical cases N, the size of the lot, is large even in comparison with n, the size of the sample. In such cases we may assume that N is infinite; we lose little in precision and gain much in simplicity. Assuming $N = \infty$, we easily find that

$$a = (1-p)^n + \binom{n}{1}p(1-p)^{n-1} + \binom{n}{2}p^2(1-p)^{n-2} + \dots + \binom{n}{c}p^c(1-p)^{n-c}.$$

We take this expression of a, the probability of acceptance, for granted and we concentrate on discussing some of its practical implications.

We graph a as function of p; see fig. 14.4. If we graphed $100a$ as function of $100p$, the form of the curve would be the same. Now, $100p$ is the percentage of defective articles in the lot submitted. On the other hand, if several lots with the same percentage of defectives were subjected to the same inspection procedure, the relative frequency of acceptance, that is, the ratio of accepted lots to submitted lots, would be close to a. Therefore, in the long run, $100a$ will be the percentage of the lots accepted among the lots submitted. This explains the labeling of the axes in fig. 14.4. The curve in fig. 14.4 allows us to survey how the procedure operates on lots of various quality, and so it is appropriately called the *operating characteristic.*

Judged by its effects, does the procedure appear reasonable? This is the question that we wish to consider.

If there are no defectives in the lot, there should be no chance for rejecting it. In fact, if $p = 0$ our formula yields $a = 1$, as it should. If there are only defectives in the lot, there should be no chance for accepting it. In fact, if $p = 1$ our formula yields $a = 0$, as it should. Both extreme points of the operating characteristic curve are obviously reasonable.

If the number of defectives increases, the chances of acceptance should diminish. In fact, differentiating with a little skill, we easily find the surprisingly simple expression

$$\frac{da}{dp} = -(n - c)\binom{n}{c}p^c(1 - p)^{n-1-c}$$

which is always negative. Therefore, the operating characteristic is necessarily a falling curve, as represented in fig. 14.4, which is again as it should be.

The absolute value of the derivative, or $-da/dp$, has also a certain practical significance. The change dp of the abscissa represents a change in the quality of the lot. The change da of the ordinate represents a change in the chances of acceptance, due to the change in quality. The larger is the ratio of these chances da/dp in absolute value, the sharper is the distinction made by the procedure between two slightly different lots. Especially, the point at which da/dp attains its maximum absolute value may be appropriately called the "point of sharpest discrimination." This point is easily recognized in the graph: it is the point of inflexion, if there is one, and otherwise the left-hand extremity of the curve. (Its abscissa is $p = c/(n - 1)$.)

(2) The rule appears sensible also from another standpoint. It has a certain flexibility. By choosing n, the size of the sample, and c, the acceptance number, we can adapt the rule to concrete requirements. Both the consumer and the producer require protection against the risks inherent in sampling. A bad lot may sometimes yield a good sample and a good lot a bad sample, and so there are two kinds of risks: the sampling procedure may accept a bad lot or reject a good lot. The consumer is against accepting bad lots and the producer is against rejecting good lots. Still, both kinds of undesirable decisions are bound to happen now and then and the only thing that we can reasonably demand is that they should not happen too often. This demand leads to concrete problems such as the following.

"Determine the sample size and the acceptance number so that there should be less than once chance in ten that a lot with 5% defectives is accepted and there should be less than five chances in a hundred that a lot with only 2% defectives is rejected."

In this problem, there are two unknowns, the sample size n and the acceptance number c. The condition of the problem requires the following two inequalities:

$$a > 0.95 \text{ when } p = 0.02,$$

$$a < 0.1 \quad \text{when } p = 0.05.$$

It is possible to satisfy these two simultaneous conditions, but it takes considerable numerical work to find the lowest sample size n and the corresponding acceptance number c for which the required inequalities hold.

We shall not discuss the numerical work. We are much more concerned here with visualizing the problem than with solving it. Let us therefore look a little further into its background. As we said already, both the acceptance of a bad lot and the rejection of a good lot are undesirable, the first from the consumer's viewpoint and the second from the producer's viewpoint. Yet the two undesirable possibilities may not be equally undesirable and the interests of consumer and producer may be not quite so sharply opposed. The acceptance of a bad lot is not quite in the interest of the producer; it may damage his reputation. Yet the rejection of a good lot may be very much against the interests of the consumer; he may need the articles urgently and the rejection may cause considerable delay. Moreover, repeated rejection of good lots, or even the danger of such rejection, may raise the price. If the interests of both parties are taken into account, the rejection of a good lot may be still less desirable than the acceptance of a bad lot. Seen against this background, it appears understandable that the conditions of our problem afford more protection against the rejection of the better quality than against the acceptance of the worse quality. (Only 5 chances in a hundred are allowed for the first undesirable event, but 10 chances in a hundred for the second.)

(3) The problem discussed under (2) admits another, somewhat different, interpretation.

The producer's lawyer affirms that there are no more than 2% defectives in the lot. Yet the consumer's lawyer contends that there are at least 5% defectives in the lot. For some reason (it may be a lot of fuzes) a full inspection is out of the question; therefore some sampling procedure has to decide between the two contentions. For this purpose the procedure outlined under (1) with the numerical data given in (2) can be appropriately used.

In fact, the conflicting contentions of the two lawyers suggest a fiction. We may pretend that there are exactly two possibilities with respect to the lot: the percentage of defectives in the lot is either exactly 2% or exactly

5%. Of course, nobody believes such a fiction, but the statistician may find it convenient: it restricts his task to a decision between two clear and simple alternatives. If the parties agree that the rejection of a lot with 2% defectives is less desirable than the acceptance of a lot with 5% defectives, the statistician may reasonably adopt the procedure outlined in (1) with the numerical data prescribed in (2). Whether the statistician's choice will satisfy the lawyers or the philosophers, I do not venture to say, but it certainly has a clear relation to the facts of the case. The statistician's rule, applied to a great number of analogous cases, accepts a good lot (with 2% defectives) about 950 times out of 1,000 and rejects it only about 50 times, but the rule rejects a bad lot (with 5% defectives) about 900 times out of 1,000 and accepts it only about 100 times. That is, the statistician's rule, which is based on sampling, cannot be expected to give the right decision each time, but it can reasonably be expected to give the right decision in an assignable percentage of cases *in the long run.*

(4) To give an adequate idea of what the statisticians are doing on the basis of just one example is, of course, a hopeless undertaking. Yet on the basis of the foregoing example we can obtain an idea of the statistician's task which, although very incomplete, is not very much distorted: the statistician designs rules of the same nature as the rule of acceptance sampling procedure outlined in (1) and considered in relation to numerical data in (2). We may understand the statistician's task if we have understood the nature of the rules he designs. Therefore, we have to formulate in general terms what seems to be essential in our particular rule; I mean the rule discussed in the foregoing sub-sections (1), (2), and (3).

Our rule prescribes a choice between two courses of action, acceptance and rejection. Yet the aspect of the problem considered under (3) is more suitable for generalization. There we considered a *choice between two statistical hypotheses.* (They were "this random sample is taken from a large lot with 2% defectives" and "this random sample is taken from a large lot with 5% defectives.") Any reasonable choice should be made with due regard to past experience and future consequences. In fact, our rule is designed with regard to both.

According to our rule, the choice depends upon a set of clearly specified observations (the testing of n articles and the number of defectives detected among the n articles tested). These observations constitute the relevant experience on which the choice is based. As our rule prefers a hypothesis to another on the basis of observations, it can claim to be named an *inductive* rule.

Our rule is designed with a view to probable consequences. The statistician cannot predict the consequences of any single application of the rule. He forecasts merely how the rule will work *in the long run.* If the choice prescribed by the rule is tried many times in such and such circumstances, it will lead to such and such result in such and such percentage of the trials, in the long run. Our rule is designed with a view to *long range consequences.*

To sum up, our *rule is designed to choose between statistical hypotheses, is based on a specified set of observations, and aims at long range consequences.* If we may regard our rule as sufficiently typical, we have an idea what the statisticians are doing: they are designing rules of this kind.

(In fact, they try to devise "best" rules of this kind. For example, they wish to render the chances of such and such undesirable effect a minimum, being given the size of the sample, on which the work and expense of the observations depend.)

(5) Taking a random sample from a lot is an important operation in statistical research. There is another problem about this operation that we have to discuss here. We keep our foregoing notation in stating the problem.

In a very large lot, $100p$ *percent of the articles is defective. In order to obtain some information about* p*, we take a sample of* n *articles from the lot, among which we find* m *defective articles. On the basis of this observation, which value should we reasonably attribute to* p*?*

There is an obvious answer, suggested by the definition of probability itself. Yet the problem is important and deserves to be examined from various angles.

Our observation yields some information about p. Especially, if m happens to be different from 0, we conclude that p is different from 0. Similarly, if m is less than n, we conclude that p is less than 1. Yet in any case p remains unknown and all values between 0 and 1 are eligible for p. If we attribute one of these values to p, we make a guess, we adopt a conjecture, we choose a statistical hypothesis.

Let us think of the consequences of our choice before we choose. If we have a value for p, we can compute the probability of the event the observation of which is our essential datum. I mean the probability for finding exactly m defective articles in a random sample of n articles. Let us call this probability P. Then

$$P = \binom{n}{m} p^m (1-p)^{n-m}.$$

The value of P depends on p, varies with p, can be greater or less. If, however, this probability P of an observed event is very small, we should reject the underlying statistical hypothesis. It would be silly to choose such an unlikely hypothesis that has to be rejected right away. Therefore let us choose the least unlikely hypothesis, the one for which the danger of rejection is least. That is, let us choose the value of p for which P is *as great as possible.*

Now, if P is a maximum, $\log P$ is also a maximum and, therefore,

$$\frac{d \log P}{dp} = \frac{m}{p} - \frac{n-m}{1-p} = 0.$$

This equation yields

$$p = \frac{m}{n}.$$

And so, after some consideration, we made the choice that we were tempted to make from the outset: as a reasonable approximation to p, the underlying probability, we choose m/n, the observed relative frequency.

Yet our consideration was not a mere detour. We can learn a lot from this consideration.

Let us begin by examining the rôle of P. This P is the probability of a certain observed event E (m defectives in a sample of size n). This probability is computed on the basis of the statistical hypothesis H_p that $100p$ is the percentage of defectives in the lot. The probability P varies with the hypothesis H_p (with the value of p). The smaller P, the less acceptable, the less likely appears H_p. Thus we are led to consider P as indicating the *likelihood* of the hypothesis H_p. This term "likelihood" has been introduced before (in sect. 7 (5)), in the same meaning, but now we may see the reasons for its introduction more clearly.

Let us emphasize that we choose among the various admissible statistical hypotheses H_p (with $0 \leqq p \leqq 1$) the one for which P, the likelihood of H_p, is as great as possible. Behind this choice there is a principle, appropriately called the *principle of maximum likelihood*, that guides the statistician also in other cases, less obvious than our case.

9. Judging non-statistical conjectures. We consider several examples in order to illustrate the same fundamental situation from several angles.

(1) The other day I made the acquaintance of a certain Mr. Morgenstern. This name is not very usual, but not unknown to me. There was a German author Morgenstern for whose nonsense poetry I have a great liking. And, Oh yes, my cousin who lives in Atlanta, Georgia, recently began work in the offices of Mark Morgenstern & Co., consulting engineers.

At the beginning I had no thoughts about Mr. Morgenstern. After a while, however, I hear that he is in the engineering business. Then other pieces of information leaked out. I hear that the first name of my new acquaintance is Mark, and that his place of business is Atlanta, Georgia. At this stage it is very difficult not to believe that this Mr. Morgenstern is the employer of my cousin. I ask Mr. Morgenstern directly and find that it is so.

This trivial little story is quite instructive. (It is based, by the way, on actual experience, but the names are changed, of course, and also some irrelevant circumstances.) That two different persons should have exactly the same last name is not improbable, provided that the name is very common such as Jones or Smith. It is more improbable that two different persons have the same first and last name, especially, when it is an uncommon name, such as Mark Morgenstern. That two different persons have the same profession, or the same large town as residence, is not improbable. Yet it is very improbable that two different persons taken at random should have the same unusual name, the same home town, and the same occupation. A chance coincidence was hard to believe and so my conjecture about my recent acquaintance Mr. Morgenstern was quite reasonable. It turned out to be correct, but this has really little to do with the merits of the case. My conjecture was reasonable, defensible, justifiable on the basis of the probabilities considered. Even if my conjecture had turned out incorrect, I would have no reason to be ashamed of it.

In this example, no numerical value was given for the probability decisively connected with the problem, but a rough estimate for it could be obtained with some trouble.

(2) Two friends who met unexpectedly decided to write a postcard to a third friend. Yet they were not quite sure about the address. Both remembered the city (it was Paris) and the street (it was Boulevard Raspail) but they were both uncertain about the number. "Wait," said one of the friends, "let us think about the number without talking, and each of us will write down the number when he thinks that he has got it." This proposal was accepted and it turned out that both remembered the same number: 79 Boulevard Raspail. They put this address on the postcard which eventually reached the third friend. The address was correct.

Yet what was the reason for adopting the number 79? By not talking to each other, the two friends made their memories work independently. They both knew that Boulevard Raspail is long enough to have buildings numbered at least up to 100. Therefore, it seems reasonable to assume that the probability for a chance coincidence of the two numbers is not superior to 1/100. Yet this probability is small, and so the hypothesis of a chance coincidence appears unlikely. Hence the confidence in the number 79.

(3) According to the statement of the bank, the balance of my checking account was $331.49 at the end of the past month. I compute my balance for the same date on the basis of my notes and find the same amount. After this agreement of the two computations I am satisfied that the amount in which they both agree is correct. Is this certain? By no means. Although both computations arrived at the same result, the result could be wrong and the agreement may be due to chance. Is that likely?

The amount, expressed in cents, is a number with five digits. If the last digit was chosen at random, it could just as well be 0 or 1 or 2, . . . or 8 as 9, and so the probability that the last digit should be 9 is just 1/10. The same is true for each of the other figures. In fact, if all figures were chosen at random, the number could be any one of the following:

000.00, 000.01, 000.02, . . . 999.99

I have here obviously 100,000 numbers. If that assemblage of five figures, 33149, was produced in some purely random way, all such assemblages could equally well arise. And, as there are 100,000 such assemblages, the probability that any one given in advance should be produced is

$$\frac{1}{100{,}000} = \left(\frac{1}{10}\right)^5 = 10^{-5}$$

Now, $10^{-5} = 0.00001$ is a very small probability. If, trying to produce an effect with such a small probability, somebody manages to succeed at the very first trial, the outcome may easily appear as miraculous. I am, however, not inclined to believe that there is anything miraculous about my modest bank account. A chance coincidence is hard to believe and so I am driven to the conclusion that the agreement of the two computations is due to the correctness of the result. Ordinary normal people generally think so in similar circumstances and after the foregoing considerations this kind of belief appears rather reasonable.

(4) To which language is English more closely related, to Hungarian or to Polish? Very little linguistic knowledge is enough to answer this question, but it is certainly more fun to obtain the answer by your own means than to accept it on the authority of some book. Here is a common sense access to the answer.

Both the form and the meaning of the words change in the course of history. We can understand the changes of form if we realize that the same language is differently pronounced in different regions, and we can understand the changes of meaning if we realize that the meaning of words is not rigidly fixed, but shifting, and changes with the context. In the second respect, however, there is one conspicuous exception: the meaning of the numerals one, two, three, . . . certainly cannot shift by imperceptible degrees. This is a good reason to suspect that the numerals do not change their meaning in the course of linguistic history. Let us, therefore, base a first comparison of the languages in question on the numerals alone. Table III lists the first ten numerals in English, Polish, Hungarian, and seven other modern European languages. Only languages that use the Roman alphabet are considered (this accounts for the absence of Russian and modern Greek). Certain diacritical marks (accents, cedillas) which are unknown in English are omitted (in Swedish, German, Polish, and Hungarian).

Looking at Table III and observing how the same numeral is spelled in different languages, we readily perceive various similarities and coincidences. The first five languages (English, Swedish, Danish, Dutch, and German) seem to be pretty similar to each other, and the next three languages (French, Spanish, and Italian) appear to be in even closer agreement; so we have two groups, one consisting of five languages, the other of three. Yet even these two groups appear to be somehow related; observe the coinciding spelling of 3 in Swedish, Danish and Italian, or that of 6 in English and French.

Table III. Numerals in ten languages.

English	Swedish	Danish	Dutch	German	French	Spanish	Italian	Polish	Hungarian
one	en	en	een	ein	un	uno	uno	jedem	egy
two	tva	to	twee	zwei	deux	dos	due	dwa	ketto
three	tre	tre	drie	drei	trois	tres	tre	trzy	harom
four	fyra	fire	vier	vier	quatre	cuatro	quattro	cztery	negy
five	fem	fem	vijf	funf	cinq	cinco	cinque	piec	ot
six	sex	seks	zes	sechs	six	seis	sei	szesc	hat
seven	sju	syv	zeven	sieben	sept	siete	sette	siedem	het
eight	atta	otte	acht	acht	huit	ocho	otto	osiem	nyolc
nine	nio	ni	negen	neun	neuf	nueve	nove	dziewiec	kilenc
ten	tio	ti	tien	zehn	dix	diez	dieci	dziesiec	tiz

Polish seems to be closer to one group in some respects, and to the other in other respects; compare 2 in Swedish and Polish, 7 in Spanish and Polish. Yet Hungarian shows no such coincidences with any of the nine other languages. These observations lead to the impression that Hungarian has little relation to the other nine languages which are all in some way related to each other. Especially, and this is the answer to our initial question, English seems to be definitely closer related to Polish than to Hungarian.

Yet there are several objections. A first objection is that "similarity" and "agreement" are vague words; we should say more precisely what we mean. This objection points in the right direction. Following its suggestion, we sacrifice a part of our evidence in order to render the remaining part more precise. We consider only the *initials* of the numerals listed in Table III. We compare two numerals expressing the same number in two different languages; we call them "concordant" if they have the same initial, and "discordant" if the initials are different. Table IV contains the number of concordant cases for each pair of languages. For instance, the

Table IV. Concordant initials of numerals in ten languages.

E	8	8	3	4	4	4	4	3	1	39
	Sw	9	5	6	4	4	4	3	2	45
		Da	4	5	4	5	5	4	2	46
			Du	5	1	1	1	0	2	22
				G	3	3	3	2	1	32
					F	8	9	5	0	38
						Sp	9	7	0	41
							I	6	0	41
								P	0	30
									H	8

number 7, in the same row as the letters "Sp" and in the same column as the letter "P" indicates that Spanish and Polish have exactly seven concordant numerals out of the possible 10 cases. The reader should check this and a few other entries of Table IV. The last column of Table IV shows how many concordant cases each language has with the other nine languages altogether. This last column shows pretty clearly the isolated position of Hungarian: it has only 8 concordant cases altogether whereas the number of concordant cases varies between 22 and 46 for the other nine languages.

Yet, perhaps, any definite conclusion from such data is rash: those coincidences of initials may be due to chance. This objection is easy to raise, but not so easy to answer. Chance could enter the picture through various channels. There may be an element of chance due to the fact that the correspondence between letters and pronunciation is by no means rigid. This is true even of a single language (especially of English). *A fortiori*, the same letter is often pretty differently pronounced in different languages and, on the other hand, different letters are sometimes very similarly pronounced. We have to admit that the coincidences observed are not free from some random element. Yet the question is: Is it *probable* that such coincidences as we have observed are due to mere chance?

If we wish to answer this question precisely, numerically, we have to adopt some precise, numerically definite statistical hypothesis and draw consequences from it which can be confronted with the observations. Yet the choice of a suitable hypothesis is not too obvious. We consider here two different statistical hypotheses.

I. There are two bags. Each bag contains 26 balls, each ball is marked with a letter of the alphabet, and different balls in the same bag are differently marked. With both hands, I draw simultaneously from both bags, one ball from each. The two letters so drawn may coincide or not; their coincidence is likened to the coincidence of the initials of the same numeral written in two different languages (and non-coincidence is likened to non-coincidence). The probability of a coincidence is 1/26.

II. The coincidence of the initials of the same numeral written in two different languages is again likened to the coincidence of two letters drawn simultaneously from two different bags and, again, both bags are filled in the same way with balls marked with letters. Yet now each of the bags contains 100 balls and each letter of the alphabet is used to mark as many different balls in the bag as there are numerals in Table III having that letter as initial. The probability of a coincidence is found to be 0.0948.

On both hypotheses, the comparison of the ten first numerals is likened to ten independent drawings of the same nature.

We can compare both hypotheses with the observations if we compute suitable probabilities. Tables V and VI contain the relevant material.

Table V compares the relative frequencies actually found with the probabilities computed. Columns (2) and (3) of Table V refer to all 45 pairs of languages considered in Table IV. Columns (4) and (5) of Table V refer only to 9 pairs, formed by Hungarian matched with the remaining nine languages. For the sake of concreteness, let us focus on the line that deals with 6 or more coincidences ($n = 6$). Such coincidences turned up in

Table V. Absolute and relative frequencies, and probabilities, for *n* or more coincidences of initials

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Frequencies				Probabilities	
n	10 languages		9 lang. v. Hu.		Hyp. II	Hyp. I
0	45	1.000	9	1.000	1.000000	1.000000
1	40	0.889	5	0.556	0.630644	0.324436
2	35	0.778	3	0.333	0.243824	0.054210
3	31	0.689	0	0 000	0.061524	0.005569
4	25	0.556	0	0.000	0.010612	0.000381
5	15	0.333	0	0.000	0.001281	0.000018
6	9	0.200	0	0.000	0.000108	0.000001
7	7	0.156	0	0.000	0.000006	0.000000

9 out of 45 cases as column (2) shows. Therefore, the observed relative frequency of 6 or more coincidences is $9/45 = 0.2$, whereas this many coincidences have only a little more than one chance in ten thousand to happen on hypothesis II, and only one chance in a million on hypothesis I; see columns (6) and (7), respectively. Similar remarks apply to the other lines of Table V: what has been actually observed appears as extremely improbable on either hypothesis, so there are strong grounds to reject both hypotheses. Yet columns (4) and (5) present a different picture: the coincidences observed are somewhat improbable on hypothesis I, but they appear as quite usual and normal from the standpoint of hypothesis II. The

Table VI. Total number of coincidences, observed and theoretical (Hypothesis II).

	Coincidences		Deviations	
	Observed	Expected	Actual	Standard
10 languages	171	42.66	128.34	7.60
9 lang. v. Hu.	8	8.53	– 0.53	2.78

impression gained from Table V is corroborated by Table VI: if we consider all 45 pairs of languages, the actually observed total number of coincidences exceeds tremendously what we have to expect on the basis of hypothesis II, yet the expected and observed numbers agree closely if we consider only the 9 pairs in which Hungarian is matched with the other 9 languages. (On hypothesis I, we have considerably stronger disagreement in both cases.)

In short there is no obvious interpretation of "chance" that would permit us to make chance responsible for all the coincidences observable in Table III; there are too many of them. Yet we can quite reasonably make chance responsible for the coincidences between Hungarian and the other languages. The explanation that Hungarian is unrelated to the other languages which are all related to each other has been vindicated.

The point is that this explanation has been vindicated, thanks to the consideration of probabilities, by so *few observations.* The explanation itself is supported by an overwhelming array of philological evidence.

(5) From appropriate observations (with telescope and spectroscope) we can conclude that certain elements found in the crust of our globe are also present in the sun and in certain stars. This conclusion is based on a physical law discovered by G. Kirchhoff almost a century ago (which says roughly that a luminous vapor absorbs precisely the same kind of light that it emits). Yet the conclusion appeals also to probabilities, and this is the point with which we are concerned here; we shall reduce the physical part of the argument to a schematic outline.

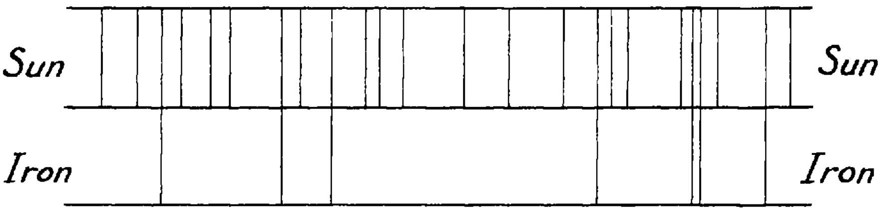


Fig. 14.5. Coincidences.

Using suitable apparatus (a prism or a diffraction grating) we can detect a sequence of lines in the light of the sun (in the solar spectrum). We can detect a sequence of lines also in the light emitted by certain substances, such as iron, vaporized at high temperature in the laboratory. (In fact, the lines in the spectrum of the sun, the Fraunhofer lines, are dark, and the lines in the spectrum of iron are bright.) Kirchhoff examined 60 iron lines and found that each of these lines coincides with some solar line. (See the rough schematic fig. 14.5 or *Encyclopaedia Britannica*, 14th edition, vol. 21, fig. 3 on plate I facing p. 560.) These coincidences are fully understandable if we assume that there is iron in the sun. (More exactly, these coincidences follow from Kirchhoff's law on emission and absorption if we assume that in the atmosphere of the sun there is iron vapor that absorbs some of the light emitted by the central part of the sun glowing at some still higher temperature.) Yet, perhaps (here is again that ever-present objection) these coincidences are due to chance.

The objection deserves serious consideration. In fact, no physical observation is absolutely precise. Two lines which we regard as coincident could be different in reality and just by chance so close to each other that, with the limited precision of our observations, we might fail to recognize their difference. We have to concede that any observed coincidence may be only an apparent coincidence and there may be, in fact, a small difference. Yet let us ask a question: Is it *probable* that each of the 60 coincidences observed springs from a random difference so small that it failed to be detected by the means of observation employed?

Kirchhoff, who registered the observed lines on an (arbitrary) centimeter scale, estimated that he could not have failed to recognize a difference that exceeded 1/2 millimeter on his scale. On this scale the average distance between two adjacent lines of the solar spectrum was about 2 millimeters. If the 60 lines of iron were thrown into this picture at random, independently from each other, what would be the probability that each falls closer to some solar line than 1/2 millimeter?

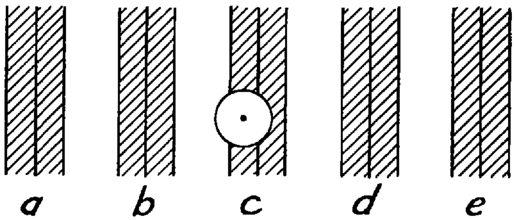


Fig. 14.6. Equidistant lines.

We bring this question nearer to its solution by formulating an equivalent question in a more familiar domain. Parallel lines are drawn on the floor; the average distance between two adjacent lines is 2 inches. We throw a coin on the floor 60 times. If the diameter of the coin is 1 inch, what is the probability that the coin covers a line each time?

In this last formulation, the question is easy to answer. Assume first that the lines on the floor are equidistant (as in fig. 14.6) so that the distance from each line to the next is 2 inches. If the coin covers a line, the center of the coin is at most at 1/2 inch distance from the line and, therefore, this center lies somewhere in a strip 1 inch wide that is bisected by the line (shaded in

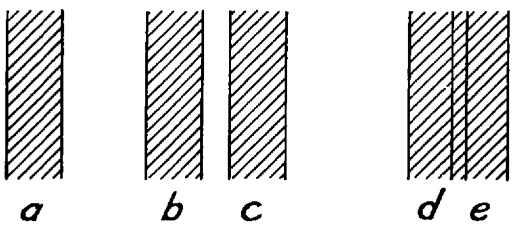


Fig. 14.7. Lines at irregular distances.

fig. 14.6). Obviously, the probability that the coin cast on the floor should cover a line is 1/2. The probability that the coin, cast on the floor 60 times, should cover some line each time, is $(1/2)^{60}$.

Assume now that the lines on the floor are not equidistant; the average distance between two adjacent lines is still supposed to be 2 inches. We imagine that the lines, which were equidistant originally, came into their present position by being shifted successively. If a line (as line *b* in fig. 14.7) is shifted so that its distance from its next neighbor remains more than 1 inch, the chances of the coin for covering some line remain unchanged. If, however, the line is so shifted (as line *d* in fig. 14.7) that its distance from the next line becomes less than 1 inch, the two (shaded) attached strips overlap and the chances of the coin to cover a line are diminished. Therefore, the required probability is less than $(1/2)^{60}$.

To sum up, if the iron lines were thrown by blind chance into the solar spectrum, the probability of the 60 coincidences observed by Kirchhoff would be less than 2^{-60} and so less than 10^{-18} or

$$\frac{1}{1{,}000{,}000{,}000{,}000{,}000{,}000}.$$

"This probability" says Kirchhoff, whom we quoted already in the motto prefixed to this chapter "is rendered still smaller by the fact that the brighter a given iron line is seen to be, the darker, as a rule, does the corresponding solar line appear. Hence this coincidence must be produced by some cause, and a cause can be assigned which affords a perfect explanation of the observed facts."

(6) The following example is not based on actual observation, but it illustrates a frequently arising, typically important situation.

An extremely dangerous disease has been treated in the same locality by two different methods which we shall distinguish as the "old treatment" and the "new treatment." Of the 9 patients who have been given the old treatment 6 died and only 3 survived, whereas of the 11 patients who received the new treatment only 2 died and 9 survived. The twenty cases are clearly displayed in Table VII.

Table VII. Four Place Correlation Table.

Patients	Died	Survived	Total
Old treatment	6	3	9
New treatment	2	9	11
Total	8	12	20

A first glance at this table may give us the impression that the observations listed speak strongly in favor of the new treatment. The relative frequency of fatal cases is

6/9 or 67% with the old treatment,

2/11 or 18% with the new treatment.

On second thoughts, however, we may wonder whether the observed numbers are large enough to give us any reasonable degree of confidence in the percentages just computed, 67% and 18%. Still, the fact remains that the number of fatal cases was much lower with the new treatment. Such a low mortality, however, could be due to chance. *How easily can chance produce such a result?*

This last question seems to be the right question. Yet, at any rate, the question must be put more precisely before it can be answered. We have to explain the precise meaning in which we used the words "chance" and "such." The word "chance" will be explained if we assimilate the present case to some suitable game of chance. A fair interpretation of the words "such a result" seems to be the following: we consider all outcomes in which the number of fatalities with the second treatment is *not higher than that actually observed.* Thus, we may be led eventually to the following formulation.

There are two players, Mr. Oldman and Mr. Newman, and 20 *cards, of which* 8 *are black and* 12 *are red. The cards are dealt so that Mr. Oldman receives* 9 *cards and Mr. Newman receives* 11 *cards. What is the probability that Mr. Newman receives* 2 *or less black cards?*

This formulation expresses as simply and as sharply as possible the contention that we have to examine: the difference between the old and the new treatment does not really matter, does not really influence the mortality, and the observed outcome is due to mere chance.

The required probability turns out to be

$$\frac{335}{8398} = 0.0399 \sim \frac{1}{25}$$

That is, an outcome that appears to be as favorable to the new treatment, as the observed outcome, or even more favorable, will be produced by chance about once in 25 trials. And so the numerical evidence for the superiority of the new treatment above the old cannot be simply dismissed, but is certainly not very strong.

In order to see clearly in these matters, let us give a moment's consideration to a situation in which the numerical data would lead us to a probability 1/10,000 instead of 1/25. Such data would make very hard to believe that the observed difference in mortality is due to mere chance but, of course, they would not prove right away the superiority of the new treatment. The data would furnish a pretty strong argument for the existence of *some non-random difference* between the two kinds of cases. What the nature of this difference actually is, the numbers cannot say. If only young or vigorous people received the new treatment, and only elderly or weak people the old treatment, the argument in favor of the medical superiority of one treatment above the other would be extremely weak.

(7) I think that the reader has noticed a certain parallelism between the six preceding examples of this section. Now this parallelism may be ripe to be brought into the open and formulated in general terms. Yet let us follow as far as possible the example of the naturalist who carefully compares the relevant details, rather than the example of those philosophers who rely mainly on words. We went into considerable detail in discussing our examples; if we do not take into account the relevant particulars carefully, our labor is lost.

In each example there is a *coincidence* and an *explanation*. (Name, surname, occupation and home town of a person I met coincide with those of a person I heard of. Explanation: the two persons are identical.—Two numbers, remembered or computed by two different persons, coincide. Explanation: the number, arrived at by two persons working independently, is correct.—The initials of several couples of numerals, designating the same number in two different languages, coincide. Explanation: the two languages are related.—The bright lines in the spectrum of iron, observed in laboratory experiments, coincide with certain dark lines in the spectrum of the sun. Explanation: there is iron vapor in the atmosphere of the sun.—A new treatment of a disease coincides with lower mortality. Explanation: the new treatment is more effective.)

Contrasting with these specific explanations, the nature of which varies with the nature of the example, there is another explanation which can be stated in the same terms in all examples: the observed coincidences are due to chance.

The specific explanations are not groundless, some of them are reasonably convincing, but none of them is logically necessary or rigidly proven. Therefore the situation is fundamentally the same in each example: there are two rival conjectures, a specific conjecture, and the "universally applicable" hypothesis of "randomness" which attributes the coincidences to chance.

Yet, if we look at it more closely, we perceive that the "hypothesis of randomness" is vague. The statement "this effect is due to chance" is ambiguous, since chance can operate according to different schemes. If we wish to obtain some more definite indication from it, we have to make the hypothesis of randomness more precise, more specific, express it in terms of probability, in short, we have to raise it to the rank of a *statistical hypothesis*.

In everyday matters we usually do not take the trouble to state a statistical hypothesis with precision or to compute its likelihood numerically. Yet we may take a first step in this direction (as in example (1)) or go even a little further (as in examples (2) or (3)). In scientific questions, however, we should clearly formulate the statistical hypothesis involved and follow it up to a numerical estimate of its likelihood, as in examples (5) and (6).

In the transition from the general and therefore somewhat diffuse idea of randomness to a specific statistical hypothesis we have to make a choice. There are cases in which we scarcely notice this choice, since we can perceive just one statistical hypothesis that is simple enough and fits the case reasonably well; in such a case the hypothesis chosen appears "natural" (as in examples (3), (5) and (6)). In other cases the choice is quite noticeable; we do not see immediately a statistical hypothesis that would be simple enough and fit the case somewhat "realistically," so we choose after more or less hesitation (as in example (4)).

Eventually there are two rival conjectures facing each other: a non-statistical, let us say "physical," conjecture *Ph* and a statistical hypothesis *St*. Now, a certain event *E* has been observed. This event *E* is related both to *Ph* and to *St*, and is so related that its happening could influence our choice between the two rival conjectures *Ph* and *St*. If the physical conjecture *Ph* is true, *E* appears as easily explicable, its happening is easily understandable. In the clearest cases (as in example (5)) *E* is implied by *Ph*, is a consequence of *Ph*. On the other hand, from the standpoint of the statistical hypothesis *St*, the event *E* appears as a "coincidence" the probability p of which can be computed on the basis of the hypothesis *St*. If the probability p of *E* turns out to be low, the happening of the event *E* is not easily explicable by "chance," that is by the statistical hypothesis *St*; this weakens our confidence in *St* and, accordingly, strengthens our confidence in *Ph*. On the other hand, if the probability p of the observed event *E* is high, *E* may appear as explicable by chance, that is, by the statistical hypothesis *St*; this strengthens somewhat our confidence in *St* and accordingly weakens our confidence in *Ph*.

It should be noticed that the foregoing is in agreement with what we said about rival conjectures in sect. 13.12 and adds some precision to the pattern of plausible reasoning discussed in sect. 12.3.

The omnipresent hypothesis of randomness is an alternative to any other kind of explanation. This seems to be deeply rooted in human nature. "Was it intention or accident?" "Is there an assignable cause or merely chance coincidence?" Some question of this kind occurs in almost every debate or deliberation, in trivial gossip and in the law courts, in everyday matters and in science.

10. Judging mathematical conjectures. We compare some examples treated in foregoing chapters with each other and with those treated in the foregoing section.

(1) Let us remember the story of a remarkable discovery told in sect. 2.6. Euler investigated the infinite series

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \cdots + \frac{1}{n^2} + \cdots.$$

First he found various transformations of this series. Then, using one of these transformations, he obtained an approximate numerical value for the sum of the series, the value 1.644934. Finally, by a novel and daring procedure, he guessed that the sum of the series is $\pi^2/6$. Euler felt himself that his procedure was daring, even objectionable, yet he had a good reason to trust his discovery: the value found by numerical computation, 1.644934, coincided, as far as it went, with the value guessed

$$\frac{\pi^2}{6} = 1.64493406 \ldots.$$

And so Euler was confident. Yet was this confidence reasonable? Such a coincidence may be due to chance.

In fact, it is not outright impossible that such a coincidence is due to chance, yet there is just one chance in ten million for such a coincidence to happen: the probability that chance, interpreted in a simple manner, should produce such a coincidence of seven decimals is 10^{-7}; cf. sect. 9 (3) and ex. 11. And so we should not blame Euler that he rejected the explanation by chance coincidence and stuck to his guess $\pi^2/6$. He proved his guess ultimately. Yet we need not insist here on the fact that it has been proved: with or without confirmation, Euler's guess was, in itself, not only brilliant but also reasonable.

(2) Let us look again at sect. 3.1 and especially at fig. 3.1 which displays nine polyhedra. For each of these polyhedra we determined F, V, and E, that is, the number of faces, vertices, and edges, respectively, and listed the numbers found in a table (Vol. I, p. 36). Then we observed a regularity: throughout the table

$$F + V = E + 2.$$

It seemed to us improbable that such a persistent regularity should be mere coincidence, and so we were led to conjecture that the relation observed in nine cases is generally true.

There is a point in this reasoning that could be made more precise: what is the probability of such a coincidence? To answer this question, we have to propose a definite statistical hypothesis. I was not able to think of one that fits the case perfectly, but there is one that has some bearing on the situation. Let me state it in setting $F - 1 = X$, $V - 1 = Y$, $E = Z$. With this change of notation, the conjectural relation obtains the form $X + Y = Z$.

We have three bags, each of which contains n balls, numbered $1, 2, 3, \ldots n$. We draw one ball from each bag and let X, Y, and Z denote the number from the first, the second, and the third bag, respectively. What is the probability that we should find the relation

$$X + Y = Z$$

between the three numbers X, Y, and Z, produced by chance?

It is understood that the three drawings are mutually independent. With this proviso the probability required is determined and we easily find that it is equal to

$$\frac{n-1}{2n^2}.$$

Let us apply this to our example. Let us focus on the moment when we succeed in verifying the hypothetical relation for a new polyhedron. For example, after the nine polyhedra that we examined initially (in sect. 3.1) we took up the case of the icosahedron (in sect. 3.2). For the icosahedron, as we found, $F = 20$, $V = 12$, $E = 30$, and so, in fact

$$(F - 1) + (V - 1) = 19 + 11 = 30 = E.$$

Is this merely a random coincidence? We apply our formula, taking $n = 30$ (we certainly could not make n less than 30) and find that such an event has the probability

$$\frac{29}{2 \times 30^2} = \frac{29}{1800} = 0.016111;$$

that is, it has a little less prospect than 1 chance in 60. We may hesitate whether we should, or should not, ascribe the verification of the conjectured relation to mere chance. Yet if we succeed in verifying it for another polyhedron, with F, V, E about as large as for the icosahedron, and we are inclined to regard the two verifications as mutually independent, we face an event (the joint verification in both cases) with a probability less than $(1/60)^2$; this event has less chance to happen than 1 in 3600 and is, therefore, even harder to explain by chance. If the verifications continue without interruption, there comes a moment, sooner or later, when we feel obliged to reject the explanation by chance.

(3) In the foregoing example we should not stress too much the numerical values of the probabilities that we computed. To realize that the probability steadily decreases as verification follows verification may be more helpful in guiding our judgment than the numerical values computed. At any rate, there are cases in which it would be hard to offer a fitting statistical hypothesis and so it is not possible to compute the probabilities involved numerically, yet the calculus of probability still yields helpful suggestions.

In sect. 4.8 we compared two conjectures concerning the sum of four squares. Let us call them conjecture A and conjecture B, respectively. Conjecture A (that we have discovered at the end of sect. 4.6) asserts a remarkable rule that precisely determines in how many ways an integer of a certain form can be represented as a sum of four odd squares. Conjecture B (Bachet's conjecture) asserts that any integer can be represented as the sum of four squares in one or more ways. Each of the two conjectures offers a prediction about the sum of four squares, but the prediction offered by A is more precise than that offered by B. Just to stress this point, let us consider for a moment a quite unbelievable assumption. Let us assume that we know from some (mysterious) source that, in a certain case, the number of representations has an equal chance to have any one of the $r + 1$ values $0, 1, 2, \ldots r$, and cannot have a value exceeding r, which is a quite large number (and this should hold both under the circumstances specified in A and under those specified in B—a rather preposterous assumption). Now, A predicts that the number of representations has a definite value; B

predicts that it is greater than 0. Therefore, the probability that A turns out to be true in that assumed case is $1/(r+1)$, whereas the probability that B turns out to be true is $r/(r+1)$. In fact, both A and B turn out to be true in that case, both conjectures are verified, and the question arises which verification yields the stronger evidence. In view of what we have just discussed, it is much more difficult to attribute the verification of A to chance, than the verification of B. By virtue of this circumstance (in accordance with all similar examples discussed in this chapter) the verification of the more precise prediction A should carry more weight than the verification of the less precise prediction B. In sect. 4.8 we arrived at the same view without any explicit consideration of probabilities.

EXAMPLES AND COMMENTS ON CHAPTER XIV

First Part

Each example in this first part begins with a reference to some section or subsection of this chapter and supplies formulas or derivations omitted in the text. The solutions require some knowledge of the calculus of probability.

1. [Sect. 3 (3)] Accept the scheme of sect. 3 (3) for representing the succession of rainy and rainless days. Say "sunny" instead of "rainless," for the sake of convenience, and let r_r, s_r, r_s, and s_s denote probabilities,

r_r for a rainy day after a rainy day,
s_r for a sunny day after a rainy day,
r_s for a rainy day after a sunny day, and
s_s for a sunny day after a sunny day.

(a) Show that $r_r - r_s = s_s - s_r$.

(b) It was said that "a rainy day follows a rainy day more easily than a rainless day." What does this mean precisely?

2. [Sect. 3 (4)] It was said that "each letter tends to be unlike the foregoing letter." What does this mean precisely?

3. [Sect. 5 (1)] Find the general expression for the numbers in column (3) of Table I.

4. [Sect. 5 (2)] Find the general expression for the numbers in column (5) of Table I.

5. [Sect. 5 (3)] (a) Find the general expression for numbers in column (3) of Table II. (b) In order to detect a systematic deviation, if there is one, examine the differences of corresponding entries (on the same row) of columns (4) and (5); list the signs.

6. [Sect. 7 (1)] If a trial consists in casting three fair dice and a success consists in casting six spots with each dice, what is the probability of n successes in n trials?

7. [Sect. 7 (2)] Among the various events reported in the story of the Reverend Galiani told in sect. 7 (1), which one constitutes the strongest argument against the hypothesis of fair dice?

8. [Sect. 7 (3)] (a) Write down the formula that leads to the numerical value $1.983 \cdot 10^{-7}$.

(b) The probability of a success is 1/3. Find the probability that 315672 trials yield precisely 315672/3 successes.

9. [Sect. 8 (1)] The expression given for a is a sum. Each term of this sum is, in fact, a probability: for what?

10. [Sect. 8 (1)] Find the abscissa of the point of inflection of the curve represented by fig. 14.4.

11. [Sect. 9 (3)] Given a number of n figures. A sequence of n figures is produced at random, perhaps by a monkey playing with the keys of an adding machine. What is the probability that the sequence so produced should coincide with the given number? [Is the answer mathematically determined?]

12. [Sect. 9 (4)] Explain the computation of the probability 0.0948.

13. [Sect. 9 (4)] Find the general expression for the numbers (a) in column (6), (b) in column (7), of Table V.

14. [Sect. 9 (4)] Explain the computation of the expected numbers of coincidences in Table VI: (a) 42.66, (b) 8.53.

15. [Sect. 9 (4)] Explain the computation of the standard deviation 2.78 in the last row and last column of Table VI.

16. [Sect. 9 (5)] Why $(1/2)^{60}$?

17. [Sect. 9 (6)] Explain the computation of the probability 0.0399. [Generalize.]

18. [Sect. 10 (2)] Derive the expression $(n - 1)/2n^2$ for the required probability.

Second Part

19. *On the concept of probability.* Sect. 2 does not define what probability "is," it merely tries to explain what probability aims at describing: the "long range" relative frequency, the "final stable" relative frequency, or the relative frequency in a "very long" series of observations. How long such a series is supposed to be, was not stated. This is an omission.

Yet such omissions are not infrequent in the sciences. Take the oldest physical science, mechanics, and the definition of velocity in non-uniform, rectilinear motion: velocity is the space described by the moving point in a certain interval of time, divided by the length of that interval, provided that the interval is "very short." How short such an interval is supposed to be is not stated.

Practically, you take the interval of time measured as short, or the statistical series observed as long, as your means of observation allow you. Theoretically you may pass to the limit. The physicists, in defining velocity, let the interval of time tend to zero. R. von Mises, in defining probability, lets the length of the statistical series tend to infinity.

20. *How not to interpret the frequency concept of probability.* The D. Tel. shook his head as he finished examining the patient. (D. Tel. means doctor of teleopathy; although strenuously opposed by the medical profession, the practice of teleopathy has been legally recognized in the fifty-third state of the union.) "You have a very serious disease," said the D. Tel. "Of ten people who have got this disease only one survives." As the patient was sufficiently scared by this information, the D. Tel. went on. "But you are lucky. You will survive, because you came to me. I have already had nine patients who all died of it."

Perhaps the D. Tel. meant it. His grandfather was a sailor whose ship was hit by a shell in a naval engagement. The sailor stuck his head through the hole torn by the shell in the hull of the ship and felt protected "because," he reasoned, "it is very improbable that a shell will hit the same spot twice."

21. An official, charged to supervise an election in a certain locality, found 30 fake registrations among the 38 that he examined the first morning. A daily paper declares that at least 99% of the registrations in that locality are correct and above suspicion. How does the daily's assertion stand up in the light of the official's observation?

22. In the window of a watchmaker's shop there are four cuckoo clocks, all going. Three clocks out of the four are less than two minutes apart: can you rely on the time that they show? There is a natural conjecture: the clocks were originally set on time, but they are not very precise (they are just cuckoo clocks) and one is out of order. If this is so, you could rely on the time shown by three. Yet there is a rival conjecture, of course: those three clocks agree by mere chance. What is the probability of such an event?

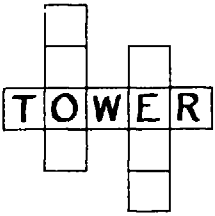
23. If a, b, c, d, e, and f are integers chosen at random, not exceeding in absolute value a given positive integer n, what is the probability that the system

$$ax + by = e$$

$$cx + dy = f$$

of two equations with two unknowns has just one solution?

24. *Probability and the solution of problems.* In a crossword puzzle one unknown word with 5 letters is crossed by two unknown words with four letters each. You guess that the unknown 5 letter word is TOWER and then you have the situation indicated by the following diagram:



In order to test your guess, you would like to find one or the other four letter word crossing the conjectural TOWER. One of the crossing words could verify the O, the other the E. Which verification would carry more weight? And why?

25. *Regular and Irregular.* Compare the two columns of numbers:

I	II
1005	1004
1033	1038
1075	1072
1106	1106
1132	1139
1179	1173
1205	1206
1231	1239
1274	1271
1301	1303

One of these two columns is "regular," the other "irregular." The regular column contains ten successive mantissas from a four-place table of common logarithms. The numbers of the irregular column agree with the corresponding numbers of the regular column in the first three digits, yet the fourth digits could be the work of an unreliable computer: they have been chosen "at random." Which is which? [Point out an orderly procedure to distinguish the regular from the irregular.]

26. *The fundamental rules of the Calculus of Probability.* In computing probabilities we may visualize the set of possible cases and see intuitively that none is privileged among them, or we may proceed according to rules. It is important for the beginner to realize that he can arrive at the same result by these two different paths. The rules are particularly important when we regard the theory of probability as a purely mathematical theory.

The rules will be important in the next chapter. For all these reasons, let us introduce here the fundamental rules of the calculus of probability, using the bag and the balls;[8] cf. sect. 3.

The bag contains p balls. Some of the balls are marked with an A, others with a B, some with both letters, some are not marked at all. (There are p possible cases and two "properties," or "events," A and B.) Let us write $\bar{A}$ for the absence of A or "non-A." (We take — as the sign of negation, but place this sign on the top of the letter, not before it.) There are four possibilities, four categories of balls.

The ball has A, but has not B. We denote this category by $A\bar{B}$ and the number of such balls by a.

The ball has B, but has not A. We denote this category by $\bar{A}B$ and the number of such balls by b.

The ball has both A and B. We denote this category by AB and the number of such balls (common to A and B) by c.

The ball has neither A nor B. We denote this category by $\bar{A}\bar{B}$ and the number of such balls (different from those having A or B) by d.

Therefore, obviously,

$$a + b + c + d = p.$$

We let $\Pr\{A\}$ stand as abbreviation for the probability of A, and $\Pr\{B\}$ for that of B. With this notation, we have obviously

$$\Pr\{A\} = \frac{a + c}{p}, \qquad \Pr\{B\} = \frac{b + c}{p}.$$

Let $\Pr\{AB\}$ stand for the "probability of A and B," that is, the probability for the joint appearance of A and B. Obviously

$$\Pr\{AB\} = \frac{c}{p}.$$

Let $\Pr\{A \text{ or } B\}$ stand for the probability of obtaining A, or B, or both A and B.[9] Obviously

$$\Pr\{A \text{ or } B\} = \frac{a + b + c}{p}.$$

---

[8] We follow H. Poincaré, *Calcul des probabilités*, p. 35–39.

[9] The little word "or" has two meanings, which are not sufficiently distinguished by the English language, or by the other modern European languages. (They are, however, somewhat distinguished in Latin.) We may use "or" "exclusively" or "inclusively." "You may go to the beach or to the movies" (not to both) is *exclusive* "or" (in Latin "aut"). "You may go the beach or have a lot of candy" is *inclusive* "or" if you mean "one or the other or both." In legal or financial documents inclusive "or" is rendered as "and/or" (in Latin "vel"). In $\Pr\{A \text{ or } B\}$ we mean the *inclusive* "or."

We readily see that

$$\Pr\{A\} + \Pr\{B\} = \Pr\{AB\} + \Pr\{A \text{ or } B\}$$

and hence follows our first fundamental rule (the "or" rule):

$$\text{(1)} \qquad \Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}.$$

We wish now to define the *conditional* probability $\Pr\{A/B\}$, in words: probability of A *if* B (granted B, posito B, on the condition B, on the hypothesis B, . . .). Also this probability is intended to represent a long range relative frequency. We draw from the bag, repeatedly, one ball each time, replacing the ball drawn before drawing the next, as described at length in sect. 2 (1). Yet we *take into account only the balls having a B.* If among the first n such balls drawn, there are m balls that also have an A, m/n is the relative frequency that should be approximately, when n is sufficiently large, equal to $\Pr\{A/B\}$. It appears rather obvious that

$$\Pr\{A/B\} = \frac{c}{b+c}.$$

In fact, there are c balls having A among the $b + c$ balls having B; also the reasoning of sect. 2 (1) may be repeated; from a certain viewpoint, we could regard the expression of $\Pr\{A/B\}$ also as a definition. At any rate, we easily find, comparing the expressions of the probabilities involved, that

$$\Pr\{A/B\} = \Pr\{AB\}/\Pr\{B\}.$$

Interchanging A and B, we find the second fundamental rule (the "and" rule):

$$\text{(2)} \qquad \Pr\{AB\} = \Pr\{A\} \Pr\{B/A\} = \Pr\{B\} \Pr\{A/B\}.$$

We can derive many other rules from (1) and (2). Observing that

$$\Pr\{A \text{ or } \bar{A}\} = 1, \qquad \Pr\{A\bar{A}\} = 0,$$

we obtain from (1), by substituting $\bar{A}$ for B, that

$$\text{(3)} \qquad \Pr\{A\} + \Pr\{\bar{A}\} = 1,$$

what we could see also directly, of course. Similarly, since

$$\Pr\{AB \text{ or } \bar{A}B\} = \Pr\{B\}, \qquad \Pr\{(AB)(\bar{A}B)\} = 0,$$

we obtain from (1), by substituting AB for A and $\bar{A}B$ for B, that

$$\text{(4)} \qquad \Pr\{B\} = \Pr\{AB\} + \Pr\{\bar{A}B\}.$$

We note here the following generalization of (2):

(5) $$\Pr\{AB/H\} = \Pr\{A/H\}\Pr\{B/HA\} = \Pr\{B/H\}\Pr\{A/HB\}.$$

We can also visualize (5) by using the bag and the balls.

27. *Independence.* We call two events independent of each other, if the happening (or not happening) of one has no influence on the chances of the other. Disregard, however, for the moment this informal definition and consider the two following formal definitions.

(I) A is called *independent of B* if

$$\Pr\{A/B\} = \Pr\{A/\bar{B}\}.$$

(II) A and B are called *mutually independent* if

$$\Pr\{A/B\} = \Pr\{A/\bar{B}\} = \Pr\{A\}, \qquad \Pr\{B/A\} = \Pr\{B/\bar{A}\} = \Pr\{B\}.$$

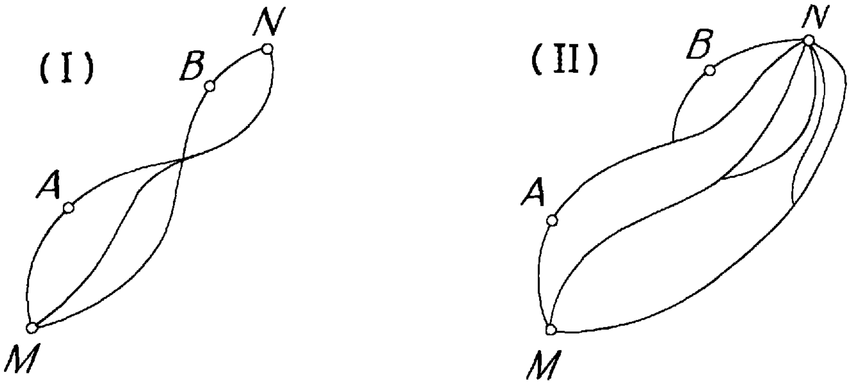


Fig. 14.8. Two systems of roads from the city M to the city N, with an essential difference.

Obviously, if A and B are mutually independent, A is independent of B. Using the rules of ex. 26, prove the theorem: *If none of the probabilities* $\Pr\{A\}$, $\Pr\{B\}$, $\Pr\{\bar{A}\}$, $\Pr\{\bar{B}\}$ *vanishes and any one of the two events A and B is independent of the other, they are mutually independent.*

28. Compare sect. 3 (5) with ex. 27.

29. A car traveling from the city M to the city N may pass through the town A and also through the town B. This is true of both systems of roads, (I) and (II), shown in fig. 14.8. Answer the following questions (a), (b), and (*c*) first in assuming that (I) represents the full system of roads between M and N, then in assuming the same thing about (II).

(a) Let A stand for the event that a car traveling from M to N passes through the town A, and B for the event that it passes through B. Assume (for both systems, (I) and (II)) that the three roads starting from M are equally well frequented (have the same probability) and also that the roads ending in N (there are 2 in (I), 6 in (II)) are equally well frequented. Find the probabilities $\Pr\{A\}$, $\Pr\{A/B\}$, $\Pr\{A/\bar{B}\}$, $\Pr\{B\}$, $\Pr\{B/A\}$, $\Pr\{B/\bar{A}\}$.

(b) Find $\Pr\{AB\}$ using the rule (2) of ex. 26.

(c) Verify that

$$\Pr\{A\} = \Pr\{B\}\Pr\{A/B\} + \Pr\{\bar{B}\}\Pr\{A/\bar{B}\},$$
$$\Pr\{B\} = \Pr\{A\}\Pr\{B/A\} + \Pr\{\bar{A}\}\Pr\{B/\bar{A}\}.$$

(d) What do you regard as the most important difference between (I) and (II)?

30. *Permutations from probability.* To decide the order in which the n participants should show their skill in an athletic contest, the name of each is written on a slip of paper, and then the n slips are drawn from a hat, one after the other, at random. What is the probability that the n names should appear in alphabetical order?

We present two solutions, and draw a conclusion from comparing them.

(1) Call E_1 the event that the slip drawn first is also alphabetically the first, E_2 the event that the slip drawn in the second place is also alphabetically the second, and so forth. The desired probability is

$$\begin{aligned}\Pr\{E_1 E_2 E_3 \ldots E_n\} &= \\ &= \Pr\{E_1\}\Pr\{E_2/E_1\}\Pr\{E_3/E_1E_2\}\ldots\Pr\{E_n/E_1\ldots E_{n-1}\} \\ &= \frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2}\cdots\frac{1}{1}.\end{aligned}$$

In fact, we obtain the first transformation by applying the rules (2) and (5) of ex. 26, and the second transformation by observing that there are n possible cases for E_1, $n-1$ for E_2 after E_1, $n-2$ for E_3 after E_1 and E_2, and so forth, whereas, for each of these events, there is just one favorable case.

(2) Call P_n the number of all the possible orderings (permutations, linear arrangements, . . .) of n distinct objects. The n names can come out from the hat in P_n ways, no one of these P_n possible cases appears as more privileged than the others, and among these P_n cases just one is favorable (the alphabetical order). Therefore, the desired probability is $1/P_n$.

(3) The results derived under (1) and (2) must be equal. Equating them, we evaluate P_n:

$$P_n = 1\cdot 2\cdot 3\ldots n = n!.$$

31. *Combinations from probability.* Mrs. Smith bought n eggs, not realizing that r of these eggs are rotten. She needs r eggs, and chooses as many among her n eggs at random. What is the probability that all r eggs chosen are rotten?

As in ex. 30 we present two solutions, and draw a conclusion from comparing them.

(1) Call E_1 the event that the first egg opened by Mrs. Smith is rotten, E_2 the event that the second egg is rotten, and so forth. The desired probability is

$$\begin{aligned}\Pr\{E_1 E_2 E_3 \dots E_r\} &= \Pr\{E_1\} \Pr\{E_2/E_1\} \Pr\{E_3/E_1E_2\} \dots \Pr\{E_r/E_1 \dots E_{r-1}\} \\ &= \frac{r}{n} \cdot \frac{r-1}{n-1} \cdot \frac{r-2}{n-2} \cdots \frac{1}{n-r+1}.\end{aligned}$$

The first transformation is obtained by rules (2) and (5) of ex. 26, the second from the consideration of possible and favorable cases for E_1, for E_2 after E_1, and so on.

(2) We have a set of n distinct objects. Any r objects chosen among these n objects form a subset of size r of the given set of size n: call C_r^n the number of all such subsets. (Usually C_r^n is called the number of "combinations" of r things selected from among n things.) In the case of Mrs. Smith's eggs, there are C_r^n possible cases, no one more privileged than the others, and among these C_r^n cases just one is favorable (if getting rotten eggs is "favorable"). Hence the desired probability is $1/C_r^n$.

(3) Comparing (1) and (2), we evaluate C_r^n:

$$C_r^n = \frac{n(n-1)\dots(n-r+1)}{1 \cdot 2 \quad \dots \quad r} = \frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

32. *The choice of a rival statistical conjecture: an example.* One person withdrew \$875 from his savings account on a certain date, and another person received \$875 two days later. The coincidence of these two amounts, one withdrawn, the other received, may be regarded as circumstantial evidence, as an indication that a crime has been committed; cf. ex. 13.6. If the jury finds it too hard to believe that this coincidence is due to mere chance, a conviction may result. Hence the problem: what is the probability of such a coincidence? The less the probability is, the more difficult it is to attribute the coincidence to chance, and the stronger is the case against the defendants.

Yet we cannot compute a probability numerically without assuming some definite statistical hypothesis. Which hypothesis should we assume? In a serious case we should give serious thought to such a question. Let us survey a few possibilities.

(1) As the number 875 has three digits, we may regard the positive integers with not more than three digits as admissible, and we may regard them as equally admissible. The probability that two such integers, chosen at random, independently from each other, should coincide, is obviously

1/999. This probability is pretty small—but is the assumption that underlies its computation reasonable?

(2) As 875 has less than five digits, we could regard all positive integers with less than five digits as equally admissible. This leads to the probability 1/9999 for the coincidence. This probability is very small indeed, but our assumption is far-fetched, even frivolous.

(3) If the point appears as important, the court can order inspection of the books of the bank or summon one of its competent officials to testify. And so it has been ascertained that immediately before the withdrawal of that sum \$875 the amount \$2581.48 was deposited on the account. In possession of this relevant information we may regard as possible and equally admissible cases the sums 1, 2, 3, . . . 2581 that could have been withdrawn from the account. Just one of these cases, 875, has to be termed favorable and so we are led to the probability 1/2581 for the coincidence. This is a small probability, but our assumption may seem reasonably realistic.

(4) We could have considered not only withdrawals in dollars, but also withdrawals in dollars and cents, such as \$875.31. If we consider all such cases as equally admissible, the probability for a coincidence becomes 1/258148. This is a very small probability, but our assumption may appear less realistic: withdrawals in dollars and cents such as \$875.31 are more usual from a checking account than from a savings account.

(5) On the contrary, one could argue that the amounts withdrawn from a savings account are usually "round" amounts, divisible by 100, or 50, or 25. Now, 875 is divisible by 25. If we regard only multiples of 25 as admissible, and equally admissible, the probability in question becomes 1/103.

Of course, we could imagine still other and more complicated ways to compute the probability, but we should not insist unduly on such a transparent example. The example served its purpose if the reader can see by now the following two points.

(a) Although some of the five assumptions discussed may seem more acceptable than others, no one is conspicuously superior to the others, and there is little hope to find an assumption that would be satisfactory in every respect and could be regarded as the best.

(b) Each of the five assumptions considered attributes a rather small probability to the coincidence actually observed, and so, on the whole, our consideration upholds the common sense view: "It is hard to believe that this coincidence is due to mere chance."

33. *The choice of a rival statistical conjecture: general remarks.* Let us try to learn something more general from the particular example considered (ex. 32). Let us reconsider the general situation discussed in sect. 14.9 (7). An event E has occurred and has been observed. Concerning this event, there are two rival conjectures facing each other: a "physical" conjecture P, and a statistical hypothesis H. If we accept the physical conjecture P, E is easily and not unreasonably explicable. If we accept the statistical hypothesis H, we can compute the probability p for the happening of such an event as E. If p is "small," we may be induced to reject the statistical hypothesis H. At any rate, the smallness of p weakens our confidence in H and therefore strengthens somewhat our confidence in the rival conjecture P.

Yet ex. 32 makes us aware that the quality of the statistical hypothesis H plays a role in the described reasoning. The statistical hypothesis H may appear as unnatural, inappropriate, far-fetched, frivolous, cheap, *unreliable* from the start. Or H may appear as natural, appropriate, realistic, reasonable, *reliable* in itself.

Now p, the probability of the event E computed on the basis of the hypothesis H, may be so small that we reject H: a rival of P drops out of the race. This increases the prospects of P—but it may increase them a lot or only a little: this depends on the quality of the rival. If the statistical hypothesis H appeared to us originally as appropriate and reliable, H was a dangerous rival and its fall strengthens P appreciably. If, however, H appeared to us as inappropriate and unreliable from the start, H was a weak rival; its fall is not surprising and strengthens P very little.

Being given a clear statistical hypothesis H, the probability p of the event E is determined, and the statistician can compute it. Yet the statistician's customer, who may be a biologist, or a psychologist, or a businessman, or any other non-statistician, has to decide what this numerical value of p means in his case. He has to decide how small a p is enough to reject or weaken the statistical hypothesis H. Yet the customer is usually not even directly interested in the statistical hypothesis H: he is primarily concerned with the rival "physical" conjecture P. And he has to decide how much weight the rejection or weakening of H has in strengthening P. This latter decision obviously cannot depend on the numerical value of p alone: it certainly depends on the choice of H.

I am afraid that the statistician's customer who wishes to make use of the numerical value p furnished by the statistician, without realizing the import of the statistical hypothesis H for his problem, just deceives himself. He can scarcely realize the import of H if he does not realize that his physical conjecture P could be also confronted with statistical hypotheses different from H. Cf. ex. 15.5.