CHAPTER I

LONG RUN FREQUENCIES

The problem of the foundation of statistics is to state a set of principles which entail the validity of all correct statistical inference, and which do not imply that any fallacious inference is valid. Much statistical inference is concerned with a special kind of property, and a good deal of the foundations depends upon its definition. Since no current definition is adequate, the next several chapters will present a better one.

Among familiar examples of the crucial property, a coin and tossing device can be so made that, in the long run, the frequency with which the coin falls heads when tossed is about 3/4. Overall, in the long run, the frequency of traffic accidents on foggy nights in a great city is pretty constant. More than 95 % of a marksman's shots hit the bull's eye. No one can doubt that these frequencies, fractions, ratios, and proportions indicate physical characteristics of some parts of the world. Road safety programmes and target practice alike assume the frequencies are open to controlled experiment. If there are sceptics who insist that the frequency in the long run with which the coin falls heads is no property of anything, they have this much right on their side: the property has never been clearly defined. It is a serious conceptual problem, to define it.

The property need not be static. It is the key to many dynamic studies. In an epidemic the frequency with which citizens become infected may be a function of the number ill at the time, so that knowledge of this function would help to chart future ravages of the disease. Since the frequency is changing, we must consider frequencies over a fairly short period of time; perhaps it may even be correct to consider instantaneous frequencies but such a paradoxical conception must await further analysis.

First the property needs a name. We might speak of the ratio, proportion, fraction or percentage of heads obtained in coin tossing, but each of these words suggests a ratio within a closed class. It is important to convey the fact that whenever the coin is tossed sufficiently often, the frequency of heads is about 3/4. So we shall say, for the present, that the *long run frequency* is about 3/4. This is only a label, but perhaps a natural one. It is hardly fitting for epidemics, where there is no long run at constant frequency of infection because the frequency is always changing. Better terms will be devised presently. It is easiest to begin with cases in which it does make sense to speak of a long run. In the end, the whole of this vague notion, the long run, must be analysed away, but in the beginning a constant reminder of its obscurity will do no harm.

The long run frequency of something is a quantity. I have called long run frequency a property, as one might call density and length properties. That convenience relies on an understood ellipsis. It is not length which is a property of a bar of iron; rather a particular length, the length of the bar, is a property of the bar. Likewise long run frequency is not a property of any part of the world, but a particular long run frequency of something may be a property of some part of the world. In what follows I shall use the word 'frequency' in connexion with this property only, and not, for example, to denote a proportion in a closed class of events.

Long run frequency is at the core of statistical reasoning. Hence the forthcoming analysis will continually employ the discoveries of statisticians, who are the people who have thought most about it. Some statisticians use the word 'probability' as the name of the physical property I have in mind, and never use that word in any other way. Others sometimes so use it and sometimes not; a few never use it to name a property. But however you think the word ought to be used, there is no denying that some statisticians do use it as the name of the property I label long run frequency. In what follows I shall not so use the word; in fact I shall scarcely use the word 'probability' at all. There is nothing wrong with using it to name a physical property if one makes plain what one is doing, but I avoid this practice to circumvent mere verbal controversy.

Most statisticians and probability theorists have no qualms over calling long run frequency a property. One of the most eminent says,

> Let the *frequency* of an outcome A in n repeated trials be the ratio n_A/n of the number n_A of occurrences of A to the total number n of trials. If, in repeating a trial a large number of times, the observed frequencies of any one of its outcomes A cluster about some number, the trial is then said to be *random*. For example, in a game of dice (two homogeneous ones)

'double-six' occurs about once in 36 times, that is, its observed frequencies cluster about 1/36. The number 1/36 is a permanent numerical property of 'double-six' under the conditions of the game, and the observed frequencies are to be thought of as measurements of the property. This is analogous to stating that, say, a bar at a fixed temperature has a permanent numerical property called its 'length' about which the measurements cluster.†

The author states that only recently have men studied this property. If someone notices a new property like this, a definition might not be needed. Perhaps even in rigorous work it might suffice to point out the property by examples and vague description, to name it, and proceed at once to investigate empirical laws in which it occurs. This will do only if the property is readily recognized, is formally much like other properties, and seems to have no peculiarities of its own. But frequency in the long run is very peculiar.

The analogy between length and frequency in the long run seems strained. Suppose, for example, that a bar is measured by a meter stick, and that the measurements cluster about 25 cm. Any measurement over 50 cm. is either wildly wrong, or was made under extraordinary conditions. Within current working theories there is no alternative. But if one measurement of double-six is 2/36, while the average is 1/36, there seems no reason to suspect error or changing conditions. There might be good reason if 2/36 were based on a very long sequence of trials. There's the well-known rub. How long? 'Many successive trials', and a 'large number of times' are all too vague. So we demand a fuller definition of this alleged physical property, long run frequency.

It might be said here that the analogy between length and frequency in the long run does not break down; there are measurements on dice tossing just as indicative of error as measurements of length. As is well known, the theory of measurements is nowadays a part of the study of frequencies. But a definition or at least deeper analysis is required before it is evident that frequency in the long run is a well-behaved property at all. Only after its analysis can it be glibly compared to more familiar things.

† Michel Loève, *Probability Theory* (New York, 1955), p. 5.

Definition

A definition of frequency is needed. The definition of a word of English is primarily a matter of linguistics, while to define an arbitrarily introduced new term may be pure stipulation. My task is neither of these, but to define a property; to draw, as it were, a line around the property to which the examples point. When, on the basis of experiment, can one conclude that something has this property? When are such hypotheses refuted by evidence, and when well supported by it?

A definition might be a form of words equivalent in meaning to the name of the property one has in mind. But that almost certainly can't be given here. There is another excellent way of defining; to state several facts about this property by way of postulates. Such a definition is for practical purposes complete if every correct inference essentially involving this property is validated by the postulates, without recourse to any unstated facts about the property—while at the same time no incorrect inference is authorized.

Such a postulational definition of frequency in the long run must not only cover what have been called direct inferences: inferences from the fact that something has a frequency of x to something else having a frequency of y. It must also cover what have been given the rather odd name of inverse inferences: from experimental data not mentioning frequency in the long run, to frequency in the long run. It has sometimes been suggested that a complete definition of the property need not cover the second kind of inference. This is absurd; a definition which does not authorize any inferring from experimental data to good support for the proposition that something has, or has not, the property in question, cannot be called a definition of an empirically significant property at all.

At one time many practical consequences might have followed a complete postulational definition. There need not be many today. The property is pretty well understood by those who use it. Their descriptions of it seldom seem completely apt, but their work with it includes some of the most remarkable discoveries made in any science. Hence our task resides squarely in the philosophy of science: to understand this great work rather than to improve it.

Statistical results will be used throughout the rest of this essay, which is little more than a philosophical codification of things already known.

Frequency in an infinite sequence

One definition of frequency in the long run is especially celebrated. 'The fundamental conception which the reader has to fix in his mind as clearly as possible is', according to Venn, 'that of a series'; he continues by describing a special kind of series which 'combines individual irregularity with aggregate regularity'.† At least part of his book may be construed as analysing a property of this kind of series. It gave expression to an idea which was much in the air at the time. Cournot, Ellis and others had written in the same vein a good deal earlier.

Von Mises refined the idea; his work justly remains the classic exposition of the theory. He studied a kind of series satisfying certain conditions, and which he called a collective. Von Mises' conditions have been modified; the modified conditions have been proved consistent. For any kind of event E, and collective K, consider the proportion of times in which events of this kind occur in the first n members of K. Call this $P_n(E)$. If $P_n(E)$ approaches a limit as n becomes large without bound, represent the limit by $P(E)$. This is a property of the collective K. Presumably it is a property of some part of the world that it would generate a collective with certain $P(E)$ for some E. $P(E)$ is von Mises' explication of frequency in the long run.

Is this property $P(E)$ empirically significant? Could the proposition that something has the property $P(E)$ in collective K ever be supported by experimental data? The questions have been asked repeatedly since Mises first published. Suppose it were granted or assumed that an experimental set-up would generate a collective. Could any data indicate the $P(E)$ in the collective? Presumably the outcome of a long sequence of actual experiments should be considered an initial or early segment of the collective. But any finite segment is compatible with, and does not give any shred of indication of, any limiting value whatsoever. On the basis of data about a finite segment, the hypothesis that a limiting value is 9/10 is, as far as experimental support is concerned, on a par with the

† John Venn, *The Logic of Chance* (London and Cambridge, 1866), p. 4.

claim that the limit is 1/10, or the suggestion that there is no limit at all. Other principles are needed than mere analyses of limits.

This standard objection is not just that no propositions about the $P(E)$ in K are conclusively verifiable or refutable in the light of experience. The trouble is that no amount of experience can, in the literal terms of the theory, give any indication whatever of the limiting value. Nor, if 'limiting value' be taken literally, is there any reason for saying observed proportions even approach a limiting value.

Von Mises' theory could be supplemented by extra principles showing how data might support, vindicate, or refute an hypothesis about the limit. But these would not be optional extras: without them there is no experimentally significant property at all. Of course von Mises' theory can be supplemented in this way. Reichenbach is one of several who attempts it. But I shall argue that the simplest principles which are adequate are also adequate for defining frequency even if frequency is not conceived as the limit of an infinite series. Infinity is redundant.

Probably von Mises did not intend his theory to be taken quite as I imply. He wished to present a mathematical conception which was in some way an idealization of the property which I label frequency in the long run. But then there remains the problem of defining that which he has idealized. Nor is it obvious that von Mises' idealization is very useful. It is sometimes said that the Euclidean plane or spherical geometry used in surveying involves an idealization. Perhaps this means that surveyors take a measurement, make simplifying assumptions, use Euclid for computing their consequences, and finally assume that these consequences are also, approximately, properties of the world from which the original measurements were taken.

It is true that some of the laws of von Mises' collective apply to frequency in the long run, and that these laws are used in computing new frequencies from old. But it is the laws, and not the infinite collective, which are of use here. Never, in the journals, will one find a statistician using the peculiar characteristics of a collective in making a statistical inference, whereas surveyors really do use some of the attributes peculiar to Euclidean plane or spherical geometry. So whatever its interest in its own right, the theory of collectives seems redundant as an idealization in the study of frequency in the long run.

Von Mises may feel that what is required from a philosophical point of view is not a mere statement of laws used by statisticians, but a coherent idealization which knits together all their principles. He might admit that you can get along without the idealization, but deny that you can see the relationship between the principles without using it. However sound may be this idea, von Mises' own theory fails to satisfy it. For the crucial principles in statistics concern the measurement of our physical property. There must be some principles for inferring frequencies from experimental data: and they are the ones which make frequency interesting. An idealization which lacks these principles fails where it is most needed.

There is no denying the intuitive appeal of replacing frequency in the long run by frequency in an infinite run. This old idea has stimulated much statistical imagination. For several decades of this century Fisher was the most prolific genius of theoretical statistics. In much of his work he refers to properties of 'hypothetical infinite populations'. These populations are at only one remove from von Mises' collectives; von Mises' work can even be construed as assigning them logical precision. However much they have been a help, I shall argue that hypothetical infinite populations only hinder full understanding of the very property von Mises and Fisher did so much to elucidate.

Axiomatic models

Most modern workers in theoretical statistics differ from von Mises. Their theory may be expressed as follows. There are various phenomena, and a property, which can be sketched by examples and vague description. The phenomena are 'random', the property, 'frequency in the long run'. The business of the formal theory of probability is to give a mathematical model of these phenomena. 'Probability' is the name given to anything satisfying the axioms of this model. Aside from some intuitive words of explanation, that is all that needs to be said about these phenomena.

This attitude is perfectly correct if one is concerned with what is now generally called probability theory, which makes deductions from a few axioms, regardless of their interpretation. But it is more dubious in statistics, which uses a unique interpretation of those axioms. Insistence on certain definable mathematical properties is of course salutary. The axioms, in a form due to Kolmogoroff, are indeed an essential tool of statistics. Frequency in the long run does, as we shall show, satisfy them: they are part of the postulational definition of frequency in the long run.

The formal theory based on Kolmogoroff's axioms is the means for proving conditional propositions, 'if the frequency in the long run of something is *this*, then that of something else is *that*'. The theory is also a superbly rich abstract discipline, one of the half dozen most stimulating fields of pure mathematics today. But pointing out examples and presenting the formal theory does not provide a complete definition of frequency in the long run. It only provides results which will assist the definition. From no theorem of the formal theory can one infer that any hypothesis about frequency in the long run is, for instance, well supported by certain experimental data. Other postulates are needed; until they are given it is not evident that any experimental property is well defined. The forthcoming work will add other postulates to those of Kolmogoroff, in the hope of completing a definition of our property.

Text-books persistently repeat the idea that the formal theory conveyed by Kolmogoroff's axioms is a model for frequency in the long run. If this does not conceal an actual confusion, it at least makes it easy to ignore problems about long run frequency. A formal mathematical theory consists essentially of vocabulary, syntax, and a set of axioms and rules of proof, all of which may be embedded in some commonly accepted logic. A theory may be modelled in another, but that does not concern us here. A science, or part of a science, is a model for a formal mathematical theory when there is a mapping between sentences of the theory, and propositions germane to the science, such that those sentences of the theory which can be derived from the axioms map on to true propositions from the science. Conversely, the theory is a model of the science when truths of the science map on to the theorems of the theory. The assertion that one is a model of the other demands, as far as I can see, that the science have clear-cut intelligible propositions. To put the case as boldly as possible, the science of long run frequency has not been proved to have such propositions. More cautiously, no clear-cut account has ever been given of their meaning or logic of verification. To call the formal theory a model of the science, and to say a term in that theory models the property of frequency in the long run, is to beg the questions at issue: are there clear-cut propositions of the science? And, how is the property to be defined?

Von Mises is plain on this very issue. What is called the formal theory of probability is a part of set theory, being the study of a special class of measures on sets of points. Referring to the property which I call frequency in the long run, von Mises insists that the theory of it 'can never become a part of the theory of sets. It remains a natural science, a theory of certain observable phenomena, which we have idealized in the concept of a collective'.

Another kind of model

To avoid confusion, it is worth mentioning a quite unexceptionable use of the word 'model' in studying frequency. It is often helpful and sometimes essential to make radical simplifying assumptions about the structure of a process. The whole of economics uses such assumptions all the time, and it is said to make models, in this sense, of coal-consumption in underdeveloped countries or of soap-production in developed ones.

There are many well-known frequency models of, for example, the transmission of a contagious disease. Suppose that every new infection increases the frequency with which further infection occurs. A model of this situation consists of an urn with *b* black and *r* red balls. Balls are drawn and replaced; if a red ball is drawn (a new infection) *c* more red balls are added also (increasing the frequency with which red balls are drawn). The urn is shaken after each replacement.† The urn provides a model of an epidemic, and so do the mathematical laws governing the behaviour of the urn. But no one would suggest that without further postulates one could ever know, for instance, how good a simplification one had made. Such models are a rich source of discoveries. But they do not have much to do with the very definition of the property they employ.

† The idea is due to G. Polya; see William Feller, *An Introduction to Probability Theory and its Applications* (New York, 1950), p. 109.

Braithwaite's theory

There is one philosophical theory nearer than any other to my ideal of postulational definition of frequency in the long run. Braithwaite proposes to state the meaning of assertions about this property in terms of rules for rejecting them on the basis of experimental data.† If these rules be added to Kolmogoroff's axioms, one has a more stringent set of postulates than the axioms alone, and which do bear on the experimental application of the property. I am certain his rules will not do, and will argue this in due course. For the present a milder remark will suffice: they are certainly not complete. For it will typically happen that on any experimental data whatsoever, a large class of statistical hypotheses will not be rejected by Braithwaite's rules. Now on the same data it would generally be admitted that some hypotheses are better supported than others. That they are better supported presumably follows, in part, from the logic of our property, frequency in the long run. No definition from which this does not follow can be wholly adequate. It does not follow from Braithwaite's rules, taken together with Kolmogoroff's axioms.

Chance

It has been convenient, in this introductory chapter, to speak of a property and label it with the phrase, 'frequency in the long run'. But already it is apparent that this term will not do. 'Frequency in the long run' is all very well, but it is a property of the coin and tossing device, not only that, in the long run, heads fall more often than tails, but also that this would happen even if in fact the device were dismantled and the coin melted. This is a dispositional property of the coin: what the long run frequency is or would be or would have been. Popper calls it a propensity of the coin, device, and situation. Now if a wine glass would break, or would have broken, or will break, when dropped, we say the glass is fragile. There is a word for the active event, and another for the passive dispositional property. It will be convenient to have a plainly dispositional word for our property—a brief way of saying what the long run frequency is or was or would have been. 'Probability' is often so used, but I eschew it here. So I shall resurrect a good

† R. B. Braithwaite, *Scientific Explanation* (Cambridge, 1953), ch. VI.