

## Statistics—Servant of All Sciences

Jerzy Neyman

The purpose of this article is to report on the first part of the Third Berkeley Symposium on Mathematical Statistics and Probability, which was held 27–30 Dec. 1954. The symposia are organized by the Statistical Laboratory, University of California, at approximately 5-year intervals; the first two were held in 1945–46 and in 1950, respectively. In organizing these symposia, the laboratory places emphasis on the *Proceedings*, which are published by the University of California Press. Although ordinary research papers are gladly accepted for publication in the *Proceedings*, preference is given to articles that tend to integrate research efforts in particular domains and that outline prospects for the future.

The *Proceedings* of the first two Berkeley symposia (1, 2) cover developments in the theory of statistics, in probability, and in various fields of application, from the physical sciences through biology and economics, to certain aspects of engineering. Owing to the generous support of the National Science Foundation, the Air Research and Development Command, the U.S. Air Force, the Office of Naval Research of the U.S. Navy, and the Office of Ordnance Research of the U.S. Army, the present symposium proceedings will be more comprehensive than the first two. It is hoped that the proceedings, five volumes of 200–300 pages each, will be published early in 1956.

The December meeting, which was held in conjunction with the annual meeting of the AAAS, represented the first part of the Third Berkeley Symposium. Because of the general character of the AAAS meetings, emphasis was

placed on applications. Thus statistics was treated mainly as a faithful “servant” of empirical sciences. However, in the view of the Statistical Laboratory, both statistics and probability are independent mathematical disciplines, and, unless they are diligently cultivated as such, their “services” can be of only limited utility. For this reason, the second part of the third symposium emphasized theory. It was held this summer as a leisurely seminar, extending through July and August.

Before proceeding to the description of the nine sessions of the third symposium that were held last December, it may be well to explain the point of view of the Statistical Laboratory on the essence of both statistics and probability and on their role in the present development of science in general.

The development of modern science is marked by a pronounced tendency toward indeterminism. A somewhat brutal description of this tendency may be stated as follows. In relation to some phenomena, instead of trying to establish a (deterministic) functional relationship between a variable  $y$ , and some other variables  $x_1, x_2, \dots, x_n$ , we try to build a (stochastic or probabilistic) model of these phenomena, predicting frequencies with which, in specified conditions, the same variable  $y$  will assume all of its possible values. For example, future research might conceivably develop methods of obtaining individual characteristics of progeny as a single-valued function of characteristics of the ancestors and of some other now unsuspected data. This would be a deterministic approach to problems of heredity. However, current genetic studies approach these problems differently and are concerned with establishing *how frequently* a progeny of a given parentage exhibits a given set of

characteristics. Similarly, rather than seeking a deterministic formula that connects with some specified variables the exact time of disintegration of a given atom, we try to develop formulas that determine *the proportions* of such atoms that will disintegrate within the next minute, within the next 2 min, and so forth.

The problem of deducing the relative frequencies (probabilities) of some events from given relative frequencies of some other events is the problem of the theory of probability. This explains the remarkable recent development of this discipline and the unusually broad range of its applications. However, the same applications require the solutions of some problems also expressed in terms of probabilities (relative frequencies), but falling outside of the usual domain of the theory of probability proper. These are the problems of statistics. They may be exemplified as follows.

Suppose a stochastic model  $M$  is advanced to represent (or to explain) certain phenomena  $P$ . The question immediately arises whether the representation is satisfactory. The observations provide a certain number, frequently only a small number, of data,  $x_1, x_2, \dots, x_n$ , and these data must be used to decide whether to accept or to reject the model  $M$ . In situations of this kind the possibility of erroneous decisions cannot be eliminated, and the best one can do is to seek methods of making decisions that, in a sense, minimize the risk of mistakes. The search for methods of making decisions on data subject to (as we call it) random variation is the subject of modern statistical theory. A rough-and-ready means of distinguishing between a problem of pure probability and a problem of statistics is to examine the quaesitum of the given problem. If the quaesitum is a probability (relative frequency), then we are in the domain of probability theory. If the quaesitum is a method of proceeding or of deciding that minimizes the probability of an error (or that minimizes a risk function defined in probabilistic terms, and so forth), then we deal with a statistical problem.

Because of these particular domains of study, the field of application of statistics and probability is literally limitless. Briefly, their role in scientific research consists (i) in providing tentative stochastic models of given classes of phe-

Dr. Neyman is chairman of the Department of Statistics and director of the Statistical Laboratory, University of California, Berkeley.

nomena and (ii) in developing optimal methods of dealing with the observations in order either to supplement the original models with certain details such as the estimates of the various constants they involve, or to decide for or against the further retention of these models. Details of my views on this subject may be found elsewhere (3).

The December 1954 program of the Third Berkeley Symposium included one session each on biology, statistical mechanics, medicine and public health, probability and induction, theory of statistics, industrial research, and psychology and two whole sessions on astronomy. Thirty-two papers were presented at these nine sessions. Under these circumstances, even a substantial report on the meeting must either be reduced to an enumeration of names and titles or be restricted to selected items. I have adopted the second alternative. Every such selection is unavoidably arbitrary and subjective. The selection adopted here is motivated partly by the desire to meet the interests of the prospective reader of this article, presumably a non-mathematician, and partly to illustrate some of the points mentioned in the introduction. In addition, the selection is somewhat influenced by the availability, at the time of writing, of the manuscripts of papers submitted for publication in the proceedings.

### Struggle for Existence and Evolution

Three papers given at the first session of the symposium were concerned with the struggle for existence and evolution. One, by J. Neyman, Thomas Park, and E. L. Scott, dealt specifically with competition of species; the two others, one by Everett R. Dempster and the other by James F. Crow and Motoo Kimura, dealt with population genetics.

The Neyman-Park-Scott paper summarized the results of several years' cooperative study conducted by the Hull Zoological Laboratory of the University of Chicago on the one hand and by the Statistical Laboratory of the University of California on the other. The starting point of these studies was the set of remarkable facts discovered by Park concerning the competition between two species of *Tribolium*, *T. confusum* and *T. castaneum*. If husbanded in isolation from each other in appropriate fixed conditions of temperature and humidity, these two species were observed to maintain reproducing populations extending over a number of years. The numbers of adult individuals of these populations fluctuated around certain points of apparent equilibrium; neither population showed any tendency toward extinction.

Denote by  $\epsilon_1$  and  $\epsilon_2$  the points of equilibrium of the two species corresponding to some fixed conditions of temperature and humidity. When these conditions change,  $\epsilon_1$  and  $\epsilon_2$  also change. We will consider the particular conditions  $C$  in which  $\epsilon_1 > \epsilon_2$  so that *T. confusum* appears more "virile" than *T. castaneum*.

In parallel with husbanding the two species of *Tribolium* separately, Park had a number of identical containers in which he attempted to develop, under the same conditions  $C$ , mixed populations of *T. confusum* and *T. castaneum*. A priori one might expect that the mixed populations will develop at some equilibrium level intermediate between  $\epsilon_1$  and  $\epsilon_2$ , with the majority of adults belonging to the more virile species. However, in a substantial number of replicates observed in six different systems of conditions  $C$ , this kind of result was never observed. Instead, invariably, one of the two competing species completely died out, leaving the field to the other, which then proceeded to establish its usual point of equilibrium. The particularly interesting points of these experiments are (i) that, under the fixed conditions  $C$ , the identity of the surviving species is not always the same, and (ii) that the apparently more virile species is not always the more frequent winner. Thus, for example, the conditions  $C$  exist in which  $\epsilon_1 > \epsilon_2$ , so that *T. confusum* appears more virile than *T. castaneum*, and yet the *T. castaneum* is predominantly the winner of the competition.

Confronted with these experimental results, one is tempted to assume that they indicate either that the active forms of one species fight (literally) with those of the other (this has never been observed) or that the contacts between the two species result in some biological changes in the particular individuals. Perhaps the females alter the frequency of laying eggs. In order to describe this kind of interaction, the term *biological interaction* was introduced. However, it is not impossible that the observed phenomena might be a consequence of another kind of interaction, termed *statistical*, that does not involve any biological changes in the individuals. Actually, one of the contributions of the Statistical Laboratory is a model of statistical interaction based on the assumption that the two competing species differ (i) in the number of eggs laid by a female in a lifetime, (ii) in the duration of inactive phases of the life-cycle (egg and pupa), and (iii) in the "voracity" of active forms (larva and adult) which eat indiscriminately the eggs (and occasionally pupae) of either species. Qualitatively, the consequences of this model agree with the observations. Unfortunately, in order to obtain workable formulas, it was neces-

sary to incorporate into the model a number of simplifying assumptions, including the hypothesis that the successive generations of *Tribolium* follow each other in a discontinuous manner, without overlaps. Whether or not the model is sufficient to represent the phenomena quantitatively will be an open question for some time. An effort to study this problem led to new experimentation and to several probabilistic studies concerned with the random walk of beetles.

Of the two papers on genetics, only one is now available, E. R. Dempster's. Its subject may be symbolized by the names of R. A. Fisher (4), Sewall Wright (5), K. Mather (6), Oscar Kempthorne (7), C. Cockerham (8), I. M. Lerner (9), and, of course, Dempster himself. It is concerned with statistical methods designed to detect the *epistatic deviations*. In order to explain the term *epistatic deviation*, we will consider a measurable characteristic  $X$  of an organism and assume for a moment that, apart from environmental variation, the value of  $X$  is determined by two pairs of genes  $A$ ,  $a$  and  $B$ ,  $b$  only. Each pair of these genes gives three different genotypes  $AA$ ,  $Aa$ ,  $aa$  and  $BB$ ,  $Bb$ ,  $bb$ , which furnish altogether nine different combinations. Denote by  $X_{ij}$ ,  $i, j = 1, 2, 3$  the average dimension of the characteristic  $X$  corresponding to the genotype symbolized by the two subscripts  $i$  and  $j$ . Thus, for example, symbol  $X_{11}$  will correspond to the genotype  $AABB$ , symbol  $X_{23}$  to the genotype  $Aabb$ , and so forth. Now, let us use a dot in order to symbolize the averaging for a given index. Thus, for example,  $X_{2.}$  will represent the average of  $X_{21}$ ,  $X_{22}$  and  $X_{23}$ . Also the symbol  $X_{..}$  will represent the grand average of all the nine different values of  $X_{ij}$ .

The values of  $X_{1.}$ ,  $X_{2.}$ , and  $X_{3.}$ , or their deviations from the grand average  $X_{..}$ , represent the population effect (or, more precisely, the average population effect) of the genes of the first pair,  $Aa$ . The same applies to averages  $X_{.1}$ ,  $X_{.2}$ , and  $X_{.3}$  in relation to the second pair of genes. Now it may happen that, for each pair of values of  $i$  and  $j$ , the genotype value  $X_{ij}$  is exactly equal to the additive combination of the average genotype effects of the two pairs of genes, so that

$$X_{ij} - X_{..} = (X_{i.} - X_{..}) + (X_{.j} - X_{..})$$

However, the exact occurrence of this equality is not very likely and the differences between the two sides of this equation are labeled *epistatic deviations*. The epistatic deviations will not equal zero if the effects of genes of one of the two pairs depend on the genetic composition of the organism with respect to the other pair. For the sake of brevity, we have considered the simplest case in which the characteristic  $X$  depends only

on two pairs of genes. In reality, the situation is much more complex, and a great number of different epistatic deviations come under consideration.

## Medicine and Public Health

Four papers were presented at the session on medicine and public health. Two, by William F. Taylor and A. T. Bharucha-Reid, dealt with the concept and with the methods of statistical evaluation of contagion. One paper, by Chin L. Chiang, Joseph L. Hodges, Jr., and Jacob Yerushalmy, was concerned with the evaluation of diagnostic tests. The last paper, by Jerome Cornfield, had for its subject the statistical problems arising from the *retrospective studies*, exemplified by the recently publicized studies of the effect of smoking on the incidence of cancer of the lung.

Although the fact is not mentioned in his paper, Taylor presented a detailed account of his own organizing-research activity as head of the biometry division, School of Aviation Medicine, U.S. Air Force. Actually, the problems sketched must represent only a section of this activity, that concerned with the concept of contagion—in accidents on the one hand and in disease on the other hand. The paper contains a well-designed background description of previous work, which was performed mostly in England and Scandinavia, and a comprehensive list of the problems facing the School of Aviation Medicine. One group of these problems—problems concerned with epidemics—was farmed out to a project at Columbia University under Bharucha-Reid; the other—problems concerned with accident proneness and accident contagion—was sent to a research project at the Statistical Laboratory, Berkeley.

Problems of the first category may be exemplified by the following. Various army camps are sometimes affected by epidemics, occasionally severe epidemics. The soldiers in these camps are housed in barracks, which, naturally, vary in a number of respects from one camp to the next. The problem arises, how much of a given epidemic is picked up outside the barracks and how much of it is the result of conditions within the barracks. A proper solution of this problem requires the development of a realistic model of what is called "within-family contagion." Some work on this subject had already been done at the Statistical Laboratory, notably by William R. Gaffey. However, his model is somewhat primitive, for it is based on the assumptions of indefinite infectiveness and zero incubation period. Reid's paper suggests that its author is well on his way toward very useful generalizations.

Among the many accident-proneness problems, the most interesting appears to be that of "tapering contagion." The first stochastic model of contagion was constructed about a quarter of a century ago by Pólya (10). For some time it was thought that the distributions this model generates are indistinguishable from those that would be observed in the absence of contagion of accidents, provided that the individuals subjected to observation vary in a particular way in their inherent proneness. Recent work (11) has shown that, if the observations are sufficiently detailed in respect to times of accidents, this difficulty does not arise. In fact, a test for the presence of the Pólya contagion has been successfully developed (12). However, closer analysis of Pólya contagion indicated that it is not likely to affect the accidents in flying. One of the basic assumptions of this theory is that, apart from a possible gain in experience, the effect of past accidents on the individual to whom these accidents occurred is independent of the time elapsed. On the other hand, intuition suggests that the effect of an accident, if any, is likely to be felt strongly in the period immediately following this accident but will probably taper off as time goes on. An initial study of tapering contagion is in progress at the Statistical Laboratory, particularly by L. M. LeCam.

In 1947 Yerushalmy (13) made a remarkable discovery that, if a radiologist makes several *independent* examinations of the same film representing an x-ray picture of an individual's lungs, then, ordinarily, the outcomes of such examinations are subject to variation: occasionally this outcome will be positive and occasionally negative. As one might expect, there are exceptions to this rule: if an individual is heavily affected by advanced tuberculosis, then all the repeated readings of the x-ray pictures by the same radiologist and by several different radiologists will be unanimously positive. However, with incipient tuberculosis and with individuals entirely free from tuberculosis, a radiologist's opinions about the same x-ray picture vary from one independent examination to the next, and this appears to be true for radiologists of highest repute. This fact itself should not be surprising because, after all, the radiologists are human!

Since this work of Yerushalmy, the problem of assessing medical diagnostic tests has been the subject of intermittent studies at the Statistical Laboratory (14, 15).

In relation to diagnostic tests with only two possible verdicts, positive or negative, the statistical approach to the assessment leads to the consideration of the probability, say  $p$ , that a single applica-

tion of the given test to a specified individual will lead to the verdict positive. At least two different values of this probability must be considered, one, say  $p_1$ , corresponding to individuals who are really ill with the disease contemplated, and the other, say  $p_2$ , corresponding to those who are free from this disease. Ideally,  $p_1$  should be equal to unity and  $p_2$  to zero. However, ideals are unattainable and one must be content with diagnostic tests for which  $p_1$  is large and  $p_2$  is small or, at least, for which  $p_1$  is substantially greater than  $p_2$ . If these numbers are known, then it is easy to arrange that repeated application of the same test leads to an arbitrarily sharp discrimination between the sick and the well. Thus, in order to assess statistically a given diagnostic test, one must face the problem of using some available or obtainable experimental data in order to estimate  $p_1$  and  $p_2$ . This problem is of particular interest in those frequent cases in which, as is the case with syphilis, there are considerable difficulties in establishing whether a given individual is or is not really affected by the given condition. In fact, the decision in this question is frequently based on the results of the same test that is being assessed. The idea of circumventing the apparently unavoidable vicious circle seems to be due to Hugo Muench (16), as follows.

The statistical approach to the problem and, in particular, the consideration of the probability  $p$  of positive response to the test, presupposes the hypothesis that, if the same test is applied several, say  $n$ , times to the same individual, with reasonable care to insure independence (with some tests this may be difficult or even impossible), then the number  $x$  of positive outcomes of the test will be a binomial variable with unknown value of the parameter  $p$  and with known exponent  $n$ . Thus, if a substantial group of persons is subjected to  $n$  independent applications of the same diagnostic test, the resulting values of the variable  $x$  may be used to estimate the distribution of  $p$  in the population. This in turn may lead to the estimates of the proportion of this population subject to the condition activating the test and to the evaluation of the effectiveness of the test.

The main difficulty with the application of this method consists in obtaining the results of repeated independent applications of the same test to a substantial number of individuals. Thus, for example, Yerushalmy's data, on which the initial three publications are based, do not satisfy the conditions indicated. The analysis was based on readings of the same x-ray pictures by five *different* physicians. The same data indicated that these physicians differed in their attitudes, and thus their readings of the

same picture cannot be really considered as *replicates* of the same experiment. Neither would it be appropriate to consider as such the readings by the same physician of four x-ray photographs of the same chest made by four widely different methods, one on the standard 14-by 17-in. Celluloid film, another on a 35-mm photofluorogram, and so forth. As a result, the afore-described papers have only methodological and illustrative significance.

The same lack of data on repeated independent applications of the same diagnostic test affects the paper presented at the symposium. One of its subjects is the analysis of the situation when a single application of the test to be assessed is followed by application, only to the individuals found positive, of a crucial test, supposed to be absolutely reliable. It is shown that, even under the simplest assumption that the population is divided into two categories, the ill and the well, without any gradation in the degree of illness, it is impossible to estimate all the parameters characterizing the situation. Another interesting subject discussed is the possibility of a sequential application of the same test to the same individuals. Namely, it appears plausible that, although a quantitative response to a test of "healthy" individuals may be very variable, with the variation overlapping that of the individuals affected by illness, the differences between responses of the same individuals to the same tests applied at two different epochs may be more stable and thus may be more adaptable for purposes of discrimination. Here, then, we arrive at the same problem discussed before, that of the distribution of the results of the same test repeatedly applied to the same individual.

In order to illustrate the distinction between "prospective" and "retrospective" studies in medicine, we will follow Cornfield and consider the much-discussed question whether or not cigarette smoking increases the chances of development of cancer of the lung.

The prospective approach would require taking under observation a substantial group of smokers and an equally substantial matching group of nonsmokers and following them for a number of years. As a result of this observation, the probabilities of contracting cancer of the lung and of surviving specified numbers of years could be estimated, and then these probabilities could serve to assess the role of smoking in the development of cancer of the lung. In particular, the data of prospective studies are sufficient for estimating correctly the *relative risk* of smokers to contract cancer of the lung compared with that of nonsmokers. This relative risk, say  $R_1$ , is measured by the quotient (proportion of smokers contracting disease)/(proportion of non-

smokers contracting the disease). If  $R_1 > 1$ , then smokers contract lung cancer more frequently than nonsmokers.

This brief description must be sufficient to indicate many of the inconveniences of the prospective method: its application must take a substantial amount of time and is certain to be very costly. The retrospective method is much easier. In application to the cancer-cigarette smoking problem, it consists in finding in hospitals a substantial group of cancer patients, say  $n_1$  and in determining among them the number  $x_1$  of those who smoke cigarettes. Then this first group is matched by another, of  $n_2$  persons, who do not suffer from cancer. Among them the number  $x_2$  of those who are smokers is counted. If the general incidence of cancer of the lung in the whole population studied is low, then the four numbers  $n_1$ ,  $x_1$ ,  $n_2$  and  $x_2$  are sufficient to obtain an approximate estimate of what may be called the *apparent* relative risk, say  $R_2$ , of cancer of the lung of smokers compared with nonsmokers. I italicize the word *apparent* in the definition of  $R_2$  in order to emphasize the difference between  $R_2$  and the formerly defined symbol  $R_1$ . The apparent risk  $R_2$  is defined as the quotient (proportion of the now living smokers who suffer from cancer of the lung)/(proportion of the now living nonsmokers who suffer from cancer of the lung).

As is true in all statistical studies, the estimation of the apparent risk from the data of a retrospective study must be subject to a sampling error. Cornfield's problem was to deduce formulas characterizing the precision of the estimate. The application of these formulas will, then, reduce the frequency of errors in judgments about  $R_2$  to a level that may be chosen in advance. This is all that a statistician can do regarding the data of retrospective studies. However, it is of some interest to mention the weakness of the method, about which little can be done, except by an equivalent of a long-drawn-out and costly prospective study.

The defect of retrospective studies consists in the fact that they can lead only to estimates of an apparent relative risk  $R_2$  but not to the relative risk  $R_1$  of contracting the disease; it is  $R_1$  rather than  $R_2$  that is of primary interest. If  $R_2$  is greater than unity then it is certainly indicative that smoking may affect the incidence of cancer of the lung, so that  $R_1$  is probably also greater than unity. However, it is essential to remember that there is no logical necessity for  $R_1$  to be greater than unity whenever  $R_2$  is greater than unity. Thus, although repeated studies indicate that cases of cancer of the lung are more frequent among the living smokers than among the living nonsmokers, this situation is perfectly consistent with the (somewhat implausi-

Table 1. Probabilities of contracting lung cancer and dying from it (fictitious).

State	Smokers	Non-smokers
$S_0$ (alive, no cancer)	0.90	0.80
$S_1$ (alive, cancer)	0.09	0.01
$S_2$ (dead from cancer)	0.01	0.19

ble) hypothesis that from the point of view of cancer of the lung smoking is helpful rather than harmful.

In order to illustrate this point we refer to the following purely fictitious situation. In order to avoid the entanglements with computing risks of death (3) from various causes, we will assume that, by some magic, all these risks can be eliminated, with the exception of cancer of the lung. Consider a person who at time  $t=0$  is alive and free from cancer of the lung. At time  $t=T$  this person may be in one of the following three states:  $S_0$ , alive and healthy;  $S_1$ , alive but suffering from lung cancer; and  $S_2$ , dead from lung cancer. Now we shall assume the probabilities of these three states at time  $T$  separately for smokers and for nonsmokers and will select these probabilities in a way that makes smoking appear very beneficial (Table 1).

It will be seen that, with these probabilities, smoking is indeed very beneficial. During time  $T$  only 10 percent of smokers get cancer and 90 percent of those who do get it survive up to the expiration of time  $T$ . On the other hand, among the nonsmokers, 20 percent contract cancer during the same period of time and, of those who do contract this disease, only 5 percent survive.

A prospective study would reveal all these details and would lead to the estimate of  $R_1 = 1/2$ . Now let us examine the possible outcome of retrospective study. For the sake of simplicity, assume that this study is to be conducted at time  $T$  in a new community set up at time  $t=0$ , which at that time was composed of 10,000 smokers and 10,000 nonsmokers, with not a single one affected by cancer of the lung. At time  $T$  the community will contain about 9900 smokers and 8100 nonsmokers. Among the smokers there will be about 900 suffering from lung cancer and among the nonsmokers there will be only 100 suffering from lung cancer. Thus, apart from sampling fluctuations, the apparent relative risk  $R_2$  of lung cancer of smokers compared with nonsmokers will be

$$R_2 = \frac{900}{9900} \div \frac{100}{8100} = \frac{81}{11} = 7.37$$

This value of the apparent risk is comparable to those actually observed and reported by Cornfield. It will be observed that the conclusions it suggests are in

striking contradiction to those indicated by the value of  $R_1 = 1/2$ , which represents the relative risk of contracting cancer of the lung.

As mentioned previously, the figures just given are purely fictitious. The sole purpose of publishing them is to call attention to the difficulties of interpreting the results of retrospective studies (17). In addition, it may be useful to mention that, under certain conditions, a special kind of study is possible that may be labeled "prospective in retrospect," which could give results essentially equivalent to those of prospective studies. For the possibility of prospective studies in retrospect, it is necessary to have an organization that keeps, as a matter of routine, detailed health records of a large number of individuals. If this is done for a substantial number of years, then, in order to investigate in retrospect the effect of any given condition  $S$  (such as smoking) on the incidence of some specified disease, it is sufficient to select from the accumulated records those referring to individuals who, say 10 years previously, had the condition  $S$  and to compare them with a similar control group. In this country data of this kind may perhaps be found in the records of the armed services and of the Veterans Administration. In Great Britain similar possibilities may exist at the National Health Service.

## Psychology

Three papers comprised the session on psychology. Frederick Mosteller discussed stochastic models of the process of learning; Herbert Solomon considered a number of statistical problems in psychometric work; and T. W. Anderson and Herman Rubin discussed statistical inference in factor analysis. Only the first of these papers lends itself to a brief description.

Mosteller's paper may be considered as a summary of the ideas crystallized in the course of preparation of a book (18) written jointly with R. R. Bush. It dealt with experiments on learning by a number of organisms, including human beings (19), rats (20), dogs (21), and paradise fish (22).

In the simplest form, an experiment consists in giving an organism a signal and then letting it make a choice between certain two actions  $A_1$  and  $A_2$ . One of these actions is "right" in the sense that, with the probability  $\pi > 1/2$ , it is followed either by a "reward" or, at least, by the avoidance of a "punishment" (for example, an electric shock). The experiments indicate that, after a number of exposures to the aforementioned trials, the organisms studied learn to associate the signal with the subsequent

reward (or punishment), not only when the connection between the two is a permanent one, so that  $\pi = 1$ , but also when  $1/2 < \pi < 1$ .

In order to treat these situations stochastically, Mosteller describes a model as follows. It is postulated that to every  $n$ th trial there corresponds a probability  $p_n$  that the experimental animal will take the right action  $A_1$ . The probability  $p_{n+1}$  of the animal's taking the same action  $A_1$  at the next trial is connected with  $p_n$  by a formula of the type

$$p_{n+1} = \alpha p_n + (1 - \alpha)\lambda$$

where  $\alpha$  and  $\lambda$  are two parameters between zero and unity, the values of which depend upon the outcome of the  $n$ th trial but not on the number  $n$ . As simple as this model appears to be, there are very considerable difficulties both in testing it against the observations and in estimating the values of the parameters it involves. Very interesting work on these subjects, both empirical and theoretical, is in progress.

## Industry

During the session on industry, three papers were presented. The first, by Albert H. Bowker, gave a broad survey of recent developments in sampling inspection, mostly developed by a large project at the department of statistics at Stanford University. The second paper, by Milton Sobel, discussed the problems of "life testing." The third, by Cuthbert Daniel, outlined various designs of industrial experimentation. All three papers are very important but are a little too technical for a brief summary in the present account.

## Astronomy

*H-R diagram.* One of the sessions on statistical problems in astronomy, with a total of five papers, was given to the H-R diagram. For the benefit of non-astronomer readers who, like myself, might think that this diagram represents a variable quantity  $H$  plotted against another variable quantity  $R$ , it may be well to explain that the two letters refer to the names of astronomers E. Hertzsprung and H. N. Russell. Quite some time past, it was suspected that the absolute brightness of stars and their temperature are connected by a relationship that depends on the chemical composition and on the size of the stars. Some 40-odd years ago, almost simultaneously, Hertzsprung and Russell had the lucky idea of plotting the estimated absolute magnitude of a number of stars against the *spectral type*, a quantity highly correlated with temperature. The resulting diagram, the H-R diagram, has a striking appearance that

reveals at a glance three principal classes of stars, the stars of the "main sequence," the "giants," and the "white dwarfs," for which the relationship between the luminosity and the temperature follows a different law.

Partly owing to the difficulties and, therefore, to the inaccuracy of measurements, the original H-R diagram had a somewhat fuzzy appearance, with the points showing a considerable scatter. However, as time went on and the methods of measurement improved, the scatter diminished, which indicated a number of refinements of the original classification of stars and led to important conclusions regarding their evolution (23).

The five papers presented at this session, by Bengt Stromgren, J. L. Greenstein, Gerald E. Kron, Harold Johnson, and Olin Eggen, are too technical for a more detailed summary.

*Spatial distribution of galaxies.* The term *galaxy* is used to denote a multimillion group of stars separated from other similar groups by colossal, relatively empty spaces. On the photographs of the sky the galaxies appear as somewhat fuzzy spots, occasionally indicating spiral organization; at times they are rather difficult to distinguish from stars.

There are two fascinating questions about galaxies. One concerns the distribution of galaxies in space. Are they distributed randomly with a sort of statistical uniformity or are they elements of still more gigantic systems? The other question is connected with the remarkable phenomenon that, judging from the position of the identifiable lines in the spectra of galaxies, practically all of them appear to recede from the Milky Way. Furthermore, the more distant a galaxy appears to be, the greater its velocity of recession. The measurements of these velocities are mostly due to the efforts of Humason and Mayall of the Mount Wilson and Palomar Observatories and the Lick Observatory, respectively. The last officially announced recession velocity amounts to about one-fifth of the velocity of light.

Although the term *velocity of recession* was used freely in the previous paragraph, I wish to emphasize that, as yet, there is no unanimity among astronomers concerning the reality of the phenomenon of recession. There is no doubt that the spectra of galaxies show shifts of spectral lines. In addition, thus far there is known only one phenomenon capable of producing shifts in the spectral lines, namely, the phenomenon of motion. A velocity of a source of light toward the observer produces shifts of spectral lines toward the violet, and a velocity away from the observer produces shifts toward the red. However, it is just possible that

similar shifts of spectral lines can be produced by some other phenomenon thus far unknown, perhaps the "aging" of light. For these reasons, it appears interesting to obtain some sort of independent evidence for or against the assumption that the galaxies recede from us (hypothesis of "expanding universe").

The three papers presented at the symposium dealt, essentially, with only the afore-mentioned two questions, or, more precisely, with the methods by which these two questions could possibly be solved. Up to the early 1930's the idea prevailed that the spatial distribution of galaxies is statistically uniform (24) except for occasional clusters. At present this idea is abandoned in favor of the idea of universal clustering. Thus clusters of galaxies become objects of independent studies. The first paper, by Fritz Zwicky, gave the first extensive collection of data regarding clusters. The second paper, by J. Neyman, E. L. Scott, and C. D. Shane was a summary of results obtained in a 5-year cooperation between the Lick Observatory and the Statistical Laboratory on the problem of distribution of galaxies. The empirical part of the study was based on the collection of plates taken by Shane and Wirtanen (25), which at the present time represents the most extensive and systematic material for statistical studies of galaxies. The theoretical part of the work included formulas characterizing the distribution of images of galaxies observable on the photographic plates both when the universe is static and when it is expanding. Roughly speaking, in the case of an expanding universe, the photographic plates would contain relatively

more images of clusters with small angular dimensions than would be the case in the absence of expansion. Unfortunately, the formulas are quite complicated, and it will be some time before their numerical evaluation can throw some new light on the problem studied.

The third paper, by George C. McVittie, was closely connected with the theory developed in Neyman, Scott, and Shane's paper. If the observed shift of the spectral lines actually is caused by velocities of expansion, then one must admit that for distant galaxies these velocities are tremendous and the observable distribution of images of galaxies is likely to be affected by relativistic effects of transmission of light. Thus, McVittie's paper dealt with modifications of the original theory that appear necessary in the light of the theory of relativity.

#### References and Notes

1. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, Berkeley, 1949).
2. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, Berkeley, 1951).
3. J. Neyman, *First Course in Probability and Statistics* (Holt, New York, 1950).
4. R. A. Fisher, "The correlation between relatives on the supposition of Mendelian inheritance," *Trans. Roy. Soc. Edinburgh* 52, 399 (1918).
5. S. Wright, "The analysis of variance and the correlations between relatives with respect to deviations from an optimum," *J. Genet.* 30, 243 (1935).
6. K. Mather, *Biometrical Genetics* (Methuen, London, 1949).
7. O. Kempthorne, "The correlation between relatives in a random mating population," *Proc. Roy. Soc. B* 143, 103 (1954).
8. C. C. Cockerham, "An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present," *Genetics* 39, 859 (1954).
9. E. R. Dempster and I. M. Lerner, "Heritability of threshold characters," *ibid.* 33, 212 (1950).

10. G. Pólya, "Sur quelques points de la théorie des probabilités," *Ann. inst. Henri Poincaré* 1, 117 (1930).
11. G. E. Bates and J. Neyman, "Contributions to the theory of accident proneness: II, True or false contagion," *Univ. Calif. Publ. in Statistics* 1, No. 10, 255 (1952).
12. G. E. Bates, "Contribution to the study of contagion," *Ann. Math. Stat.*, in press.
13. J. Yerushalmy, "Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques," *Public Health Repts. U.S.* 62, (1947), 1432 (1947).
14. J. Neyman, "Outline of statistical treatment of the problem of diagnosis," *ibid.* 62, 1449 (1947).
15. C. L. Chiang, "On the design of mass medical surveys," *Human Biology* 23, No. 3, 242 (1951).
16. H. Muench, "The probability distribution of protection test results," *J. Am. Statistical Assoc.* 31, 677 (1936).
17. A referee warns me that in spite of the fictitiousness of the figures in Table 1 and in spite of the emphasis on the methodological character of my remarks, the "tobacco people" may pick up the argument and use it for publicity purposes. I most sincerely hope that this does not happen. The purpose of this note is to reemphasize that my comments were prompted by what appears to be insufficient attention given to the distinction between the prospective and the retrospective studies. As to the validity or otherwise of the assertions regarding the incidence of cancer of the lung among smokers and among nonsmokers, I have no direct information.
18. R. R. Bush and F. Mosteller, *Stochastic Models for Learning* (Wiley, New York, in press).
19. G. A. Miller and W. J. McGill, "A statistical description of verbal learning," *Psychometrika* 17, 369 (1952).
20. C. Graham and R. M. Gagne, "The acquisition, extinction, and spontaneous discovery of a conditioned operant response," *J. Exptl. Psychol.* 26, 251 (1940).
21. R. L. Solomon and L. C. Wynne, "Traumatic avoidance learning: acquisition in normal dogs," *Psychol. Monogr.* 67, No. 354 (1954).
22. T. R. Wilson and R. R. Bush, personal communication.
23. O. Struve, *Stellar Evolution* (Princeton Univ. Press, Princeton, N.J., 1950).
24. E. P. Hubble, "Effects of red shifts on the distribution of nebulae," *Astrophys. J.* 84, 517 (1936).
25. C. D. Shane and C. A. Wirtanen, "The distribution of extra-galactic nebulae," *Astronom. J.* 59, 285 (1954).

## Wendell M. Latimer, Chemist

Wendell Mitchell Latimer was one of Gilbert N. Lewis' more important selections when he was building his chemistry department at the University of California. Latimer had an important part in shaping the department and succeeded Lewis as dean of the College of Chemistry at Berkeley. Chemistry lost one of its more versatile and prolific contributors on 6 July 1955 at the age of 62. Weakened by successive gall-bladder operations, Latimer appeared to be recovering satisfactorily when a recurrence of the difficulty weakened him further

and he died in his sleep. He leaves his wife Glatha (Hatfield) Latimer, a son Robert Milton Latimer, who is a student of chemistry, a daughter Mrs. Eleanor L. Colborn and two grandchildren Diane and Robert Edgar Colborn.

Latimer was born in Garnett, Kansas, 22 April 1893, the only son of Walter and Emma Mitchell Latimer. He entered the University of Kansas, planning to become a lawyer. He found that he enjoyed mathematics and sought some subject to which he might apply it. His first contact with chemistry came during his third year

at the university. The subject captured his interest and he decided to become a chemist. He was partly self-supporting as a student and was employed to make meteorological observations. He received the B.A. degree from the University of Kansas in 1915 and served as instructor there from 1915 to 1917. Latimer came to Berkeley as a graduate student because of G. N. Lewis' reputation and his study of some of Lewis' papers. In 1919 he received the Ph.D. degree from the University of California. His research, under the direction of G. E. Gibson, was concerned with low-temperature calorimetry. He was retained as a member of the staff and attained the full professorship in 1931. He served as assistant dean of the College of Letters and Science, 1923-24, as dean of the College of Chemistry 1941-49, and chairman of the department of chemistry 1945-49. He was Guggenheim Memorial Foundation fellow in Munich in 1930. He was associate editor of the *Journal of Chemical*