Statistical Science
1991, Vol. 6, No. 4, 320–338

Abstracts of Lectures Presented at the Statistics Seminar, MIT, 1942–1943

EDITOR'S INTRODUCTION

A series of lectures held at the Massachusetts Institute of Technology's Department of Economics and Social Science in 1942-1943 on statistics included a galaxy of leading figures in statistics and related fields from that time and since. The majority of them still survive. Three are Nobel laureates: Trygve Haavelmo, Lawrence Klein and Paul Samuelson. Doubtless, others have similar stature, but were omitted from consideration by the Nobel rules that exclude mathematical fields. Two graduate students, Lawrence Klein and Joseph Ullman, seeing the importance of this burgeoning new statistics subject, particularly in wartime, and this scientific community, organized these lectures and provided the brief outlines included here. It is entirely possible that this seminar had its own influence on the thinking of the lecturers and thereby affected their own careers; for example, Haavelmo's Nobel Prize was based on the work presented at MIT.

These lectures were brought to our attention by Steve Stigler, one of the *Statistical Science* editors. Lawrence Klein has returned 48 years after organizing the conference, participating in editing this collection and in keeping contact with the surviving contributors. He writes on his reminiscences of the conference, to introduce the collection. After the abstracts, Paul Samuelson, still very active at MIT, provides his reminiscences.

The Statistics Seminar, MIT, 1942–1943

Lawrence Klein

The graduate program in economics at MIT was introduced in the academic year 1941–1942. I, with other aspiring economists, joined the program in September 1942, the second entering class. Fresh from UC Berkeley, where I had studied with Jerzy Neyman's group in statistics at the same time that I was first investigating mathematical economics, I was naturally attracted to specialization in econometrics at MIT, where Paul Samuelson and Harold Freeman were responsible for that branch of economics.

The statement of purpose of the Econometric Society, formed in 1929, is that it supports the advancement of economics through its relationship to statistics and mathematics. At MIT, in 1942, mathematical statistics was taught and researched primarily in the mathematics department, while economics (and possibly other subjects) combined their own strengths in statistics, oriented toward the substantive discipline, with general mathematical statistics coming from the mathematics department.

Joseph Ullman, now of the mathematics faculty at the University of Michigan, and I, together with some of the other graduate students in economics, felt the need for extra knowledge about mathematical statistics. Ullman and I, with modest support from our department, organized a seminar series in statistics. We combined expository and pedagogical contributions with some presentations that were based on original research.

Naturally, we drew upon people in the Cambridge area who were working in the field of mathematical statistics but went as far as New York or Washington to round out the roster of speakers. Ullman and I kept notes of the lectures and checked our summaries with the invited speakers for the purpose of preparing an annual report. An informal report of the final version was circulated to those who might be interested in our activities.

Lawrence Klein is Benjamin Franklin Professor of Economics Emeritus, University of Pennsylvania, Department of Economics, 3718 Locust Walk, Philadelphia, Pennsylvania 19104.

The seminar continued in the next academic year, but lacked the enthusiasm of 1942–1943, as many of the participants drifted away, mainly to war-related activities. It was a surprise to me, almost 50 years later, to learn that there was an interest in reprinting our summaries.

The list of seminar speakers contains an impressive array of well-known scholars, whose presentations speak for themselves, even though the summaries are quite brief. Dirk Struik and Richard von Mises spoke on foundations of probability concepts; Kenneth Arnold and Albert Bowker talked about problems that they were occupied with as members of the statistics group in the MIT Mathematics Department; Harold Freeman and Paul Samuelson represented statistical analysis from the side of the MIT Economics Department; Harold Hotelling and Abraham Wald of Columbia University were invited as outstanding pioneers in the introduction of mathematical and statistical methods in economics; and E. B. Wilson brought wisdom from one of a great many fields that he had mastered. He was a truly broad-based scholar in the classical mold, and taught a stimulating course in mathematical economics at Harvard that I and other graduate students from MIT attended.

From the viewpoint of econometrics, which was a motivating force behind the creation of the seminar series, there is at least one paper that deserves special attention, as it is an early exposition (not the first, however) of research that directly led to the awarding of a prize in economic sciences in memory of Alfred Nobel. Trygve Haavelmo's lecture, summarized in our report, presented to the seminar his recently published ideas on the simultaneous-equations approach in econometrics. His lecture was based on his celebrated paper "The Statistical Implications of a System of Simultaneous Equations." This paper was published in *Econometrica* (vol. 11, January 1943, pages 1–12). One year later, when I completed my Ph.D. at MIT, I went to work on issues generated by Haavelmo, at the Cowles Commission for Research in Economics at the University of Chicago.

The lecture of Will Feller can also be seen as an exposition of ideas that later came to fuller fruition in his very successful book *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1950).

Norbert Weiner's lecture was, as ever, stimulating and perhaps was a prelude to his interests in economics, time series analysis and automatic control developed at greater length in his provocative book on *Cybernetics* (MIT, Cambridge, Mass., 1948). In that book, he used the Ergodic Theorem in both statistical mechanics and time series analysis with information theory. His main thrust in *Cybernetics* was not directly related to economics, but he felt that the ideas generated by the subject of cybernetics would have important economic implications.

The lectures in the seminar series were all given in the spirit of supplementation of our regular academic program, not always as new contributions or findings, but as explanations of various topics or fields of mathematical statistics. In that respect, they were very successful and brought many eminent scholars into our midst. Nearly one half of the authors are now deceased. Although it seems to be surprising that this record is being re-examined, it is gratifying to those still living to learn that it remains of interest after all these years.

From the authors who are still living, the editor of this review has received thoughtful remarks about the seminar. Kenneth Arnold has even improved the statement of the summary of his lecture. Albert Bowker has responded with some remarks about his interests at the time with Latin Squares, as well as the state of statistics in Cambridge at the time of the seminar. Harold Freeman and Trygve Haavelmo both commented in correspondence on some of the great and unusual personalities who spoke to the seminar. Among others, they both referred to Norbert Wiener.

Paul Samuelson's reminiscences provide more detailed insight into the fundamental contributions of the well-known personalities who favored us with their thoughts on very academic subjects in wartime Cambridge.

TABLE OF CONTENTS

Struik, Prof. Dirk J.
The Foundations of the Theory of Probabilities

von Mises, Dr. Richard
The Probability of Occupancy

Wald, Prof. Abraham
A Problem in Multivariate Analysis

Wiener, Prof. Norbert
Ergodic Theory

Wilson, Prof. E. B.
Contingency Tables

FOREWORD

The following are abstracts of a series of lectures sponsored by the Department of Economics and Social Science at the Massachusetts Institute of Technology. We have written these abstracts on the basis of notes taken at the seminars, although in some cases the speakers have chosen to write their own abstracts or modify those which we have written.

We wish to thank the several speakers for their generous contribution of time and effort in order to appear before the seminar group. We are also grateful to Mrs. E. Clemence for handling the preparation of the manuscript.

Lawrence R. Klein
Joseph L. Ullman

SPHERICAL PROBABILITY

Dr. Kenneth J. Arnold
MIT

Most of distribution theory is confined to a flat space or to abstract space. The theory in a flat space usually depends on the metric of the space. The only system really free of this limitation is the Edgeworth system, in which the metric is determined on the basis of the probability distribution.

The normal curve has many properties by which it can be characterized. All of these properties depend on the metric of the flat space. We shall take two of these characterizations and see what happens when they are applied to a circular one-dimensional space and to a spherical two-dimensional space.[1]

One characterization of the normal curve is that it is the Green's function of the equation of heat flow,

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{\partial \psi}{\partial t},$$

which is, for a line

$$\psi = \frac{1}{\sqrt{\pi t}} e^{-x^2/(4t)}.$$

In the case of a circle, the Green's function of

$$\frac{\partial^2 \psi}{\partial \theta^2} = \frac{\partial \psi}{\partial t}$$

is

$$\psi = \frac{1}{2\pi}\left\{1 + 2\sum_{n=1}^{\infty} \cos n\theta \, e^{-n^2 t}\right\} = \frac{1}{2\pi}\vartheta_3\left(\frac{\theta}{2}, e^{-t}\right),$$

the Jacobian Theta function. As might be expected, the same result can be obtained by wrapping the normal curve around the circle and adding densities.

$$\psi = \frac{1}{\sqrt{2\pi}\,\sigma}\sum_{n=-\infty}^{\infty} e^{-(\theta+2n\pi)^2/2\sigma^2} = \frac{1}{2\pi}\vartheta_3\left(\frac{\theta}{2}, e^{-\sigma^2/2}\right).$$

It is interesting to note that if we introduce Θ as the origin of deviations and write

$$\psi = \frac{1}{2\pi}\vartheta_3\left(\frac{\theta - \Theta}{2}, q\right),$$

Θ and q are the angular and radial coordinates of the center of gravity of the distribution. This distribution has been discussed by Wintner and Zernike.

Another characterization of the normal curve is to say it is the curve for which the arithmetic mean of the observations gives in a certain sense the best measure of central tendency. To be specific, the normal curve is the function $\psi(x, m)$ that satisfies the equation

$$\sum_{i=1}^{n} \frac{\partial \log \psi(x, m)}{\partial m} = 0$$

for all sets of x's for which m is the arithmetic mean, that is, for all for which

$$\sum_{i=1}^{n} (x_i - m) = 0.$$

This equation can be modified for the circle to read

$$\sum_{i=1}^{n} \sin(\theta_i - \Theta) = 0$$

and in combination with

$$\sum_{i=1}^{n} \frac{\partial \log \psi(\theta, \Theta)}{\partial \Theta} = 0$$

[1] Dr. Arnold suggests the following change (April 24, 1991): We shall take two of these characterizations, which have been applied to a circular one-dimensional space, and see what happens when they are applied to a spherical two-dimensional space.

yields the function

$$\psi(\theta,\Theta) = \frac{e^{k\cos(\theta-\Theta)}}{2\pi I_0(k)},$$

where $I_0(k)$ is the zeroth order Bessel function of imaginary argument. It is not surprising that the maximum likelihood estimates for Θ and k are expressible in terms of the center of gravity of the observed points. Θ is the angular coordinate of the center of gravity, and the equation for the determination of k is

$$a = \frac{I_1(k)}{I_0(k)},$$

where $I_0(k)$ and $I_1(k)$ are the zeroth and first-order Bessel functions, respectively. This distribution has been discussed by von Mises.

Applying these ideas to the sphere, the Green's function of

$$\frac{\partial}{\partial x}(1-x^2)\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial t}$$

is

$$\kappa(\theta,\varphi,t) = \frac{1}{2}\sum_{n=0}^{\infty}(2n+1)P_n(x)e^{-n(n+1)t},$$

where

$$x = \cos\theta\cos\Theta + \sin\theta\sin\Theta\cos(\varphi-\Phi).$$

The center of gravity gives estimates of t, Θ and Φ.

The von Mises process applied to the sphere gives the function

$$\psi(\theta,\varphi) = \frac{k}{4\pi\sinh k}e^{kx},$$

where x is as before. The maximum likelihood estimates of Θ and Φ are the angular coordinates of the center of gravity and the radial coordinate a gives the equation for k,

$$a = \operatorname{ctnh} k - \frac{1}{k}.$$

Similar distributions are obtained for axes of a circle or sphere rather than points on its circumference or surface.[2]

For the heat flow solution, two point distributions are added. For the center of gravity solution on the circle

$$\psi(\vartheta) = \frac{e^{k|\cos(\theta-\Theta)|}}{4\int_0^{\pi/2}e^{k\cos\theta}\,d\theta},$$

and on the sphere

$$\psi(\theta,\varphi) = \frac{k}{4\pi(e^k-1)}e^{kx},$$

where again

$$x = \cos\theta\cos\Theta + \sin\theta\sin\Theta\cos(\varphi-\Phi).$$

For the sphere, k is found from the center of gravity of an appropriate hemisphere by the equation

$$\frac{e}{e^k-1} - \frac{1}{k} = a.$$

ENUMERATION OF LATIN SQUARES

Albert H. Bowker
MIT

Associate with an $n \times n$ Latin Square a set of transformations T_1, $T_2, \ldots, T_n$ such that T_i changes $1, 2, \ldots, n$ into the ith row. If T_1, $T_2, \ldots, T_n$ is a Latin Square, $ST_1, ST_2, \ldots, ST_n$ is also a Latin Square where S is any element of the permutation group on n letters.

If $T_1 = I$ (the identity transformation), the Latin Square is standard, and this property is preserved under transformations of the nature ST_iS^{-1}. T_2 is the generating element, and the number of such elements was given by Cayley.

Consider two standard Latin Squares with generating elements T_2 and $\overline{T}_2$, distinct and conjugate. It was proved that T_2 and $\overline{T}_2$ generate the same number of Latin Squares. Hence to enumerate Latin Squares it is necessary to consider one example of each conjugate set of generating elements, and enumerate the total number of possibilities in each case by trial and by assisting theorems.

The method of Norton and Fisher and Yates is one of classification by leading diagonal and seems to result in more types and subsequent exhaustive trials. For Graeco-Latin Squares, we have the problem of existence as well as the problem of enumeration. Stevens' proof of the existence of Graeco-Latin Squares on side p^n (where p is a prime number) was discussed.

STOCHASTIC PROCESSES

Prof. Will Feller
Brown University

A stochastic process is any process whose evolution we are able to follow and predict in terms of probability. The necessity of a theory of stochastic processes is made clear by the abundance of nonsensical results attained by analyzing data without considering the causal process accounting for the data.

This is especially true in correlation analysis. The correlation of z $(= ax + by)$ with x will yield

[2] Dr. Arnold suggests the following change (April 24, 1991): Similar distributions are obtained for cases in which points appear in pairs located at opposite ends of a diameter of a circle or sphere.

any results between −1 and 1 according to the values of a and b alone, even if the variables x and y are causally unrelated. The fitting of the logistic law of growth to data leads to weird predictions, when it is not first shown that the law of growth $dy/dt = ay - by^2$ (upon which the logistic law is based), is valid for the complete process.

The mathematical theory of stochastic processes is based upon the concept of chance variable. A chance variable can assume any one of a set of values, each of which has an assigned probability. A stochastic process is measured by a number registered, say, at discrete times. Each measurement is a chance variable, $y(t)$, and the stochastic process is the sequence of chance variables $\{y(t)\}$, (t usually assuming the integral values from zero to infinity).

If the probability relations of any $y(m), \ldots, y(n)$ are the same as those of $y(m+1), \ldots, y(n+1)$, the process is said to be stationary, or temporally homogeneous. The next most simple type of connection is that of a Markoff process, in which the probability distribution of $y(t+1)$, calculated under the assumption that all previous $y(i)$ have been assigned values, depends only on the value assigned to the immediately preceding value $y(t)$.

When $y(t)$ runs through all values, and the increments of $y(t)$: $y(t+h) - y(t)$ over nonoverlapping time intervals are mutually independent, the sequence $\{y(i)\}$ is called a differential process. The Poisson process is a useful illustration, although it is more usually derived as a limiting form of the binomial law.

Let $y_n(t)$ be the probability that n events occur by time t. The probability of an event occurring in a time interval h will be λh (that is to say, it is independent of t). It follows that $y_n(t+h) = y_n(t)[1 - \lambda h] + y_{n-1}(t)\lambda h + 0(h)$. This gives rise to $y_n'(t) = \lambda y_{n-1}(t) - \lambda y_n(t)$ when we take the limit as $h \to 0$. A solution to this equation is

$$y_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!},$$

the Poisson law.

An advantage of this approach is that it lends itself quite automatically to important generalizations.

Let us consider the distributions of events arising from different universes of relative weights $f_1, f_2, \ldots$, with $\lambda_1 h, \lambda_2 h, \ldots$, the probability that an event arising from the appropriate universe occurs in a time interval h. It can be shown that the number of events that will be observed by time t is given by

$$\int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!}\, dU(\lambda),$$

where U is the function associating λ_i and f_i.

The inverse problem of discovering the stochastic process that will give rise to any given law presents difficulties. An alternative set of assumptions, due to Polya, was shown to lead to the same generalized Poisson process as above, so that, from the final law alone, the corresponding assumptions could not be distinguished. According to Polya, the distribution indicates a "contagion" of probabilities. The fact that the other interpretation implies independent events shows the dangers of applied statistics and leads to the general conclusion that a purely phenomenological approach is impossible in statistics.

When trying to discover the stochastic process giving rise to a distribution law, for example the Pareto law of distribution of income, alternative processes may be discovered, and the appropriate one can then be selected by specific economic laws.

BAYES' THEOREM AND TESTING HYPOTHESES

Prof. H. A. Freeman
MIT

Industrial sampling inspection may sometimes be regarded as a problem in estimation. If, from an infinite lot of unknown fraction defective x, a sample of size n containing m defectives is drawn, Bayes' Theorem gives

$$(1)\quad P(a \le x \le b) = \frac{\int_a^b \pi(x)x^m(1-x)^{n-m}dx}{\int_0^1 \pi(x)x^m(1-x)^{n-m}dx},$$

where $\pi(x)$ is the a priori probability of x.

Inspection can also be regarded as a problem in the testing of hypotheses. If we test the hypothesis $x = u$ with alternative hypotheses $x > u$, then, at least for large n, the best critical region R is given by

$$(2)\qquad R = 1 - \sum_{r=0}^{r=g-1} C_r^n u^r(1-u)^{n-r},$$

where g, the allowable number of defectives, is fixed by the size of R. Poisson's integral form for (2) is

$$(3)\quad R = \frac{n!}{(g-1)!(n-g)!}\int_u^1 x^{g-1}(1-x)^{n-g}dx.$$

While R is demonstrably the best critical region for *each* hypothesis in the band $0 \le x \le u$ (alternatives $x > u$), we take it to be best for the entire band-hypothesis $o \le x \le u$ (alternative $x > u$). Now for the many $\pi(x)$ that can be represented by power series, (1) and (3) are easily compared by repeated use of

$$I_x(j+1, k) = I_x(j,k)\frac{\Gamma(j+k)}{\Gamma(j+1)\Gamma(k)}x^j(1-x)^k,$$

and it is at once apparent that for known $\pi(x)$ the two procedures, estimation and hypothesis-testing, will yield strikingly different results.

SOME PROBLEMS OF STATISTICAL INFERENCE ARISING IN ECONOMETRICS

Trygve Haavelmo
The Norwegian Shipping and Trade Mission

An example of an econometric problem is the following: Let u_t be consumption in period t, v_t investment, r_t income; and let us consider the following model:

$$u_t = \alpha r_t + \beta + x_t, \qquad t = 1, 2, \ldots, N.$$
$$v_t = \gamma(u_t - u_{t-1}) + y_t, \quad u_0 = \text{a given constant},$$
$$u_t + v_t = r_t,$$

where x_t and y_t are certain random variables. A method that has been widely used but that may lead to erroneous results is to estimate the parameters α, β and γ from each equation separately by the method of least squares without taking account of the interrelationships involved. Such a procedure need not lead to an unbiased estimate. The proper way to determine the parameters is to estimate them as parameters of a joint probability distribution. Assume as known the joint distribution (density function) of the $2N$ random variables $x_1, x_2, \ldots, x_N; y_1, y_2, \ldots, y_N$. Let this assumed distribution be represented by $p(x_1, x_2, \ldots, x_N; y_1, y_2, \ldots, y_N)$. Then the joint distribution of $u_1, u_2, \ldots, u_N; v_1, v_2, \ldots, v_N$ can be derived from p.

Let this distribution be $p_1 = p_1(u_1, u_2, \ldots, u_N; v_1, v_2, \ldots, v_N)$. Then $p_1 = [1 - (1 + \gamma)\alpha]^N p$. One can then, under certain conditions, use the method of maximum likelihood to estimate α, β and γ as parameters of p_1.

Another type of problem that may arise from the same example is one of testing of an hypothesis. Suppose, for example, that one wanted to test the hypothesis that α, β and γ are all positive. Select first Ω, the class of all a priori admissible hypotheses p_1, and then subclass of Ω, call it ω, which includes the restriction that α, β and γ are all positive. By using the Neyman–Pearson theory of testing of hypotheses, we can derive rules for accepting or rejecting the hypothesis, H_0, that $p_1 \varepsilon \omega$.

SOME UNSOLVED PROBLEMS OF STATISTICAL THEORY

Prof. Harold Hotelling
Columbia University

The first class of problems has to do with the foundations of the theory. The Neyman–Pearson idea of a power function opens the door to such problems. In a large class of problems we already know uniformly most powerful tests. In case the power function depends upon a second parameter, it is sometimes possible to incorporate an estimate of the "nuisance" parameter in a new statistic and arrive at a uniformly most powerful test. Some problems involving nuisance parameters are those of contingency tables and of the testing of significance of the difference between the means of two samples when the variances need not be the same. Fisher and Yates have dealt with the former problem by the use of the subpopulation of tables with the same marginal totals.

A second class of problems comes under the heading of statistical decisions. This is not the mere acceptance or rejection of an hypothesis. For example, when the null hypothesis is rejected, the question arises as to what hypothesis is actually accepted. In the use of analysis of variance, one wants to know more than that two yields, say, are significantly different. But rather it is desired to choose the best of a variety of yields. A case of much interest is that of the choice of the degree of a curve that is to be fitted to data by the method of least squares.

Recently there have been solutions of many problems in the design of experiments, and many new problems. Sequential experiments offer great possibilities since experiments can often be made more efficient when arranged in stages. Indian statisticians have done some good work in experimental design by taking into account both the efficiency desired and the cost involved in organizing a census of the jute crop. Experiments for the purpose of determining the maximum or minimum of a function, or a point of inflection, offer an important and fascinating new field of study. Here a limited number of sets of values of the independent variables are to be chosen for experimentation. The experiments yield the corresponding values of the dependent variable, with errors.

Last there are the numerical problems. Computational methods have become a topic of growing importance. Modern multivariate analysis calls for the solution of large systems of equations in which great accuracy is desired, and this clearly calls for the construction of efficient methods of calculation. Techniques based on the theory of matrices have a place here. Iterative methods have been found in this way that converge relatively fast, and yield definite and useful limits of error.

NEYMAN'S SMOOTH TEST FOR GOODNESS OF FIT

Lawrence R. Klein
MIT

The basic paper by Neyman on the smooth test, being published in a not easily accessible journal, should be brought to the attention of all statisticians who are not familiar with this work.

The chi-square test of Pearson for goodness of fit is not powerful in guarding against smooth alternatives and may well lead to the committing of errors of the second kind. The smooth test, or the ψ^2 test, is designed especially to be powerful against smooth alternatives. The hypothesis, H_0, that is to be tested is that the random variable x is distributed according to the law $P(x \mid H_0)$, where H_0 completely specifies the distribution function. By the transformation

$$y = \int_{-\infty}^{x} p(u \mid H_0)\, du,$$

we can test the hypothesis, h_0, that y is distributed in the following manner:

$$p(y \mid h_0) = 1, \quad \text{for } 0<y<1,$$
$$p(y \mid h_0) = 0, \quad \text{for all other } y.$$

Testing the hypothesis h_0 is shown to be equivalent to testing H_0. The alternative hypothesis, or the smooth alternatives are given by

$$p(y \mid \theta_1, \theta_2, \dots, \theta_k) = Ce^{\sum_{i=1}^{k} \theta_i \pi_i(y)},$$

where

$$\pi_i(y) = a_{i0} + a_{i1} y + a_{i2} y^2 + \cdots + a_{ii} y^i.$$

The polynomials $\pi_i(y)$ are orthonormal in the interval (0, 1). The hypothesis h_0 is among the alternatives, for it is

$$\theta_1 = \theta_2 = \theta_3 = \cdots = \theta_k = 0.$$

An unbiased critical region of type C is chosen so as to satisfy the following criteria.

(1) The power function must have first and second derivatives with respect to the θ_i.

(2) The power function at the point representing h_0 shall have the value ε, where $0 < \varepsilon < 1$.

(3) The first derivatives of the power function at the point $\theta_1 = \theta_2 = \cdots = \theta_k = 0$ shall vanish.

(4) The mixed second-order partial derivatives shall vanish at the point corresponding to h_0.

(5) All other second-order partial derivatives (i.e., *not* mixed) shall have a common value at $\theta_1 = \theta_2 = \cdots = \theta_k = 0$, and furthermore this common value shall be greater than or equal to that for any other region satisfying the other criteria.

For large values of n (the number of observations in the sample), the conditions for the unbiased critical region of type C are satisfied by the inequality

$$\psi_k^2 = \sum_{i=1}^{k} u_i^2 \geq \psi_\varepsilon^2(k),$$

where

$$u_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \pi_i(y_j)$$

and $\psi_\varepsilon(k)$ is determined by

$$\frac{1}{\Gamma(k/2)} \int_{\psi_\varepsilon^2/2}^{\infty} t^{(k-2)/2} e^{-t}\, dt = \varepsilon.$$

Recently various short-cut methods in the actual calculations have simplified the application of this test.

GRAM-CHARLIER SERIES

Prof. Paul A. Samuelson
MIT

Two classes of series can be considered. Type A is

$$F(x) \sim a_0\varphi(x) + a_1\varphi'(x) + \cdots + a_r\varphi^r(x) + \cdots,$$

where

$$\varphi(x) = e^{-x^2}.$$

Type B is

$$F(x) \sim a_0\psi(x) + a_1\Delta\psi(x) + a_2\Delta^2\psi(x) + \cdots + a_r\Delta^r\psi(x) + \cdots,$$

where

$$\Delta\psi(x) = \psi(x) - \psi(x-1) \quad \text{and} \quad \psi(n) = \frac{e^{-x}x^n}{n!}.$$

A method of fitting a frequency function can be given when the parent function $\varphi(x)$ satisfies only a few requirements, namely, that all moments exist and that $\varphi(x)$ have very high contact and derivatives of every order.

$$\int_{-\infty}^{\infty} \varphi(x) x^k\, dx \quad \text{must exist for all integer } k.$$

$$\lim_{x \to \pm\infty} \varphi^k(x) x^k = 0 \quad \text{for all integer } k.$$

If we define

$$L_n[\varphi^r(x)] = \int_{-\infty}^{\infty} \frac{x^n}{n!} \varphi^r(x)\, dx,$$

then it is easily shown that the following property holds:

$$L_n[\varphi^r(x)] = \begin{cases} L_{n-r}[\varphi(k)](-1)^r, & \text{when } n \geq r, \\ 0, & \text{when } n < r. \end{cases}$$

By applying the operator L_n to each side of

$$F(x) = \sum_{k=0}^{n} a_k \varphi^k(x),$$

and making use of the above property, a set of equations to determine the a's can be obtained:

$$a_0 = L_0[F(x)]$$
$$-a_1 = L_1[F(x)] - a_0 L_1[\varphi(x)]$$
$$\vdots$$
$$a_n = L_n[F(x)] - \sum_{i=0}^{n-1} (-1)^i L_{n-i}[\varphi(x)].$$

The method of Charlier was to determine a polynomial $S_n(x)$ such that

$$\int_{-\infty}^{\infty} S_i(x)\varphi^j(x)\,dx = 0$$

except when $i = j$, and then the integral is to be equal to unity. For the Type A distribution the suitable polynomials are the Hermite polynomials. The a's are determined from the formula

$$a_n = \int_{-\infty}^{\infty} S_n(x)F(x)\,dx.$$

Another method that makes use of the moment generating function can be used. Define

$$M(\alpha:\varphi) = \int_{-\infty}^{\infty} e^{\alpha x}\varphi(x)\,dx.$$

The following property holds:

$$M[\alpha:\varphi^k(x)] = (-1)^k\alpha^k M[\alpha:\varphi(x)],$$
$$F(x) \sim \sum_{i=0}^{n} a_i\varphi^i(x),$$
$$M[\alpha:F(x)] = M\left[\alpha:\sum_{i=0}^{n} a_i\varphi^i(x)\right],$$
$$M[\alpha:F(x)] = \sum_{i=0}^{n} a_i M[\alpha:\varphi^i(x)],$$
$$M[\alpha:F(x)] = M[\alpha:\varphi(x)]\left\{\sum_{i=0}^{n}(-1)^i a_i\alpha^i\right\},$$
$$\frac{M[\alpha:F(x)]}{M[\alpha:\varphi(x)]} = \sum_{i=0}^{n}(-1)^i a_i\alpha^i.$$

Corresponding results can be derived for the generalization of the Type B discrete case where differences, sums and factorial moment-generating functions take the place of derivatives, integrals and moment-generating functions.

THE FOUNDATIONS OF THE THEORY OF PROBABILITIES

Prof. Dirk J. Struik
MIT

Two main difficulties at the foundations of the theory of probability are the definition of equally likely events, and the relations between the laws of causal natural science and the laws of statistical regularity.

The first problem arises in the purely formal theory of probability, which is essentially a theory of measure. According to which events are defined as equally likely (in a continuous case, which variables are defined as independent), different solutions are obtained to the same problem, and each is correct relative to the corresponding definition.

This arbitrariness is removed when experimental facts are considered, since the material (or social) events are subjected to laws that may determine the cases (events or variables) that are equally likely. The heuristic principle of insufficient reason, which has proved useful in determining equally likely cases in some problems, should be replaced by the principle of cogent reasoning, which determines the conditions different cases have to satisfy in order to be equally likely. In problems dealing with rigid bodies, for example, the dynamics of rigid bodies should supply us with a principle for determining equally likely cases.

Buffon's "needle problem" is a classical example that illustrates this shift of emphasis. In a plane parallel lines are drawn at a distance d. A line segment of length $l < d$ is placed in an arbitrary way on the plane. What is the probability that the line segment will intersect one of the parallel lines? The answer is indeterminate, depending on the variables introduced to determine the position of the line with respect to the parallel line. But if we state the problem as the "needle problem," in which a needle is thrown on a table, by a consideration of the dynamical laws that govern the movement of the needle, we are able to discriminate between equally likely cases. The most general conclusions concerning probabilities that can be deduced from laws governing a dynamical system are embodied in ergodic theorems.

We can draw the conclusion that a probability problem has a definite sense only if the dynamics governing the system under discussion are known. "Dynamics" is used in the general sense of a well-defined set of causes that may be analyzed, and even brought into the form of a dynamics similar to Newtonian dynamics.

Thus the "dynamics" of a mortality table is the basic set of causes that accounts for the statistical stability. Although there is a particular causal sequence that leads to each death, there is a more general type of causality behind the mortality, and it is this in which an insurance company is really interested. The proof lies in the care with which they classify the general causes of death, as pneumonia, heart attack, etc. This indicates that there are certain general levels of causality that influence statistical regularity. It is important, in this respect, to notice that statistical consideration of phenomena is not a substitute for the investigation of the unique causal determination of each event, but is rather the investigation of a different, and complementary, aspect of nature. Without causality there is no statistics, but statistics can give us no more than a statistical form of causality.

We must warn against thinking that statistics is a "result" of causality, which is after all no more than the eighteenth-century idea that probability is subjective and only due to our ignorance. There is an interplay of statistical regularity and causality, the exact forms of which must be investigated in each field anew.

THE PROBABILITY OF OCCUPANCY

Dr. Richard von Mises
Harvard University

In a group of 60 persons, three individuals had their birthdays on the same day. Is this a rare event? The correct answer to this problem is not simply a statement of the probability of getting one triply occupied day, but rather the expected number of triply occupied days. This example is illustrative of the problem of occupancy. Classically, one studied the distribution of $n_1, n_2, \ldots, n_N$ where n_t is the number of persons on the tth place. The new problem is to characterize the placement by the probability $p\{x_0, x_1, \ldots, x_k\}$, where x_s is the number of places occupied by s persons each,

$$E(x_s) = n\binom{k}{s}\left(1 - \frac{1}{n}\right)^{k-s}\left(\frac{1}{n}\right)^s.$$

In the above example:

$$E(x_0) = 309.60,$$
$$E(x_1) = 51.03,$$
$$E(x_2) = 4.14,$$
$$E(x_3) = 0.22.$$

In physical problems, the question arises as to what values of x_i $(i = 1, 2, \ldots, k)$ fulfilling the conditions

$$x_0 + x_1 + \cdots + x_K = n$$
$$x_1 + 2x_2 + \cdots + Kx_K = K$$

make $P\{x_0, x_1, \ldots, x_n\}$ a maximum.

$$p\{x_0, x_1, \ldots, x_k\} = \frac{k!}{n^k}\frac{n!}{x_0! \cdots x_k!} q_o^{x_0} q_1^{x_1} \cdots q_k^{x_k},$$

where $q_s = 1/s!$. By the usual method of differentiation and use of Stirling's formula, one could maximize

$$p\{x_0, x_1, \ldots, x_k\},$$

but there is no reason to believe that the values b_i of x_i that will be found in this way will be integers, as they obviously must be. From

$$\frac{P\{y_0, y_1, \ldots, y_k\}}{P\{x_0, x_1, \ldots, x_k\}} = \frac{x_0!\, x_1! \cdots x_k!}{y_0!\, y_1! \cdots y_k!} b_0^{(y_0 - x_0)} \cdots b_k^{(y_k - x_k)}$$

follows in the case of $y_0 = x_0$; $y_1 = x_1 + 1$; $y_2 = x_2; \ldots; y_k = x_k$ that

$$\frac{P\{y_0, \ldots, y_k\}}{P\{x_0, \ldots, x_k\}} = \frac{b_1}{c_1 + 1}.$$

Thus, if the x_i were subject to no condition, $x_i = [b_i]$ would be the solution that leads to the maximum $P\{x_0, x_1, \ldots, x_k\}$, for this is the point at which

$$\frac{P\{y_0, \ldots, y_k\}}{P\{x_0, \ldots, x_k\}}$$

changes from less than unity to greater than unity. The problems of finding the integers that fulfill the two linear conditions and make P a maximum leads to the following algebraic question.

Let

$$\xi_s = x_s - [b_s],$$
$$R = \sum_{s=1}^{k} b_s - \sum_{s=1}^{k} [b_s],$$
$$S = \sum_{s=1}^{k} sb_s - \sum_{s=1}^{k} s[b_s].$$

Then the two Diophantic equations

$$\sum_{r=1}^{k} \xi_r = R$$

and

$$\sum_{r=1}^{k} r\xi_r = S$$

admit an infinity of solutions. A solution ξ_i is called a "smallest" solution, if no other solution ξ_i' exists where all ξ_i' are at least as close to zero as ξ_i. One has to determine the region in which such "smallest" solutions can fall. The answer is

$$|x_s - [b_s]| = |\xi_s| < \frac{k^2(k+1)}{2} + k|R| + |S|.$$

If the b_i have been computed in the usual way, R and S are known, and then this inequality for $x_i - [b_i]$ shows how far the correct solution x_i may deviate from $[b_i]$.

A PROBLEM IN MULTIVARIATE ANALYSIS

Prof. Abraham Wald
Columbia University

Consider a set of random variables $x_1, \ldots, x_p$ with the joint probability density function $f(x_1, \ldots, x_p)$. Suppose that nothing is known about

$f(x_1, \ldots, x_p)$ except that it is a continuous function of $x_1, \ldots, x_p$. A sample of n independent observations is drawn and the αth observation on x_i is denoted by $x_{i\alpha}$ $(i = 1, \ldots, p;\ \alpha = 1, \ldots, n)$. In quality control of manufactured products, the problem of setting tolerance limits for the variates $x_1, \ldots, x_p$ is of importance. This problem can be formulated as follows: *For some given positive values $\beta < 1$ and $\gamma < 1$ we need to construct p pairs of functions $L_i(x_{11}, \ldots, x_{pn})$ and $M_i(x_{11}, \ldots, x_{pn})$ $(i = 1, \ldots, p)$ such that the probability that*

$$\int_{L_p}^{M_p} \cdots \int_{L_1}^{M_1} f(t_1, \ldots, t_p)\, dt_1 \cdots dt_p \geq \gamma$$

holds is equal to β. The function L_i is called the lower and M_i the upper tolerance limit of x_i.

This problem has been satisfactorily solved by Wilks in the univariate case, that is, when $p = 1$. An important feature of Wilks' solution is that the probability distribution of

$$Q_1 = \int_{L_1}^{M_1} f(t_1)\, dt_1$$

does not depend on the unknown density function $f(t_1)$. A natural extension of Wilks' method leads to the difficulty that the distribution of

$$Q_p = \int_{L_p}^{M_p} \cdots \int_{L_1}^{M_1} f(t_1, \ldots, t_p)\, dt_1 \cdots dt_p$$

depends on the unknown density function $f(t_1, \ldots, t_p)$. However, by a slight modification of the construction of tolerance limits the probability distribution of Q_p can be made independent of $f(t_1, \ldots, t_p)$. The tolerance limits are defined as follows: Let the observations $x_{11}, \ldots, x_{1n}$ be arranged in order of increasing magnitude. Then $L_1 = x_{1r_1}$ and $M_1 = x_{1s_1}$, where r_1 and s_1 are some properly chosen integers. To obtain the tolerance limits L_i and M_i $(i = 2, \ldots, p)$, let S be the set of all points $q_\alpha = (x_{1\alpha}, \ldots, x_{p\alpha})$ for which $L_j < x_{j\alpha} < M_j$ $(j = 1, \ldots, i-1)$ and arrange the ith coordinates of the points in S in order of increasing magnitudes. Then L_i is equal to the r_ith element and M_i is equal to the s_ith element of this ordered sequence.

The probability distribution of Q_p is given by

$$p(Q_p)\, dQ_p = \frac{\Gamma(n+1)}{\Gamma(s_p - r_p)\Gamma(n - s_p + r_p + 1)} \cdot Q_p^{s_p - r_p - 1}(1 - Q_p)^{n - s_p + r_p} dQ_p.$$

If β, γ, r_p and s_p are given, the value of n can be obtained by solving the equation

$$\int_\gamma^1 P(Q_p)\, dQ_p = \beta.$$

A simple rectangle is not always the most efficient shape of region to use. In the case of strongly correlated variates this would lead to an unnecessarily large area. For this reason the theory is extended to the general case when the tolerance region is composed of several small rectangles.

ERGODIC THEORY

Prof. Norbert Wiener
MIT

A basic tool of ergodic theory is Lebesgue measure. The measure of a set of points on a line was used as an illustration, and defined. Using the same illustration, a measure preserving transformation was introduced. In this case, it is a one-to-one transformation of the points on an interval into the same interval, so that the Lebesgue measure is unchanged.

Probability theories have been based on the axiom that the ratio of successful events to total events approaches a constant value as the number of events observed gets very large. On the basis of the concepts introduced, a simple form of the ergodic theorem can be formulated. This theorem makes it possible to replace the *assumption* that a limit is approached, by a *proof* that under certain general conditions a limit is, in fact, approached.

Consider the point P, one of a set of points on an interval. Let $T(P)$ be the coordinate of that point after a measuring preserving transformation. In like manner $T[T(P)]$ or $T^2(P)$ is the coordinate after two such transformations. The ergodic theorem states that the proportion of the points P, $T(P)$, $T^2(P), \ldots, T^n(P), \ldots$ that falls on any specified part of the interval approaches a limit as n approaches infinity.

Let T be interpreted as transforming P from one time to another. The sequence P, $T(P)$, $T^2(P), \ldots$ is a time series, say, of prices. The theorem allows us to give a definite value to the probability that price will be observed in any specified range, subject to the qualification we will next consider.

The question arises whether the successive elements of any given time sequence can be considered to be the result of repeated applications of a measure preserving transformation. Perhaps this condition is fulfilled for meteorological time sequences because of the stability of the underlying forces. For economic time series, however, this condition is often not fulfilled and caution must be taken in applying the theorem.

CONTINGENCY TABLES

Prof. E. B. Wilson
Harvard University

In biological and medical experiments many contingency tables arise that cannot be analyzed by the chi-square test because of low cell frequencies. A method for treating such cases is presented. Although illustrated with one four-fold universe with two marginal totals fixed, besides the total number in the sample, the general principle can be stated for v cellular universes with not necessarily equal numbers of cells, and with L totals remaining fixed, including the v totals of the size of the subsample from each universe.

Consider now the sample of N from a four-fold universe having two characters A and B, with probabilities $p_1 = p_{AB}$, $p_2 = p_{\alpha B}$, $p_3 = p_{A\beta}$ and $p_4 = p_{\alpha\beta}$. If n_i is the number in a sample of N having the attribute associated with p_i, then the probability of observing n_1, n_2, n_3, n_4 is given by

$$P = \frac{N!}{n_1!n_2!n_3!n_4!} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}.$$

We restrict our further attention to those tables that have the following totals fixed: $n_1 + n_3 = (A)$, $n_1 + n_2 = (B)$ and $N = n_1 + n_2 + n_3 + n_4$, and also satisfy the condition that the probability of their occurrence shall not vary from table to table by virtue of the values of p_1, p_2, p_3 and p_4.

These conditions are sufficient to determine an associated universe that represents the appropriate null hypothesis. It can also be shown that the probability of an observed table arising from a universe satisfying the null hypothesis is

$$\left[\sum \frac{1}{n_1!n_2!n_3!n_4!}\right]^{-1} \frac{1}{n_1!n_2!n_3!n_4!},$$

where the summation is over all samples which could arise, satisfying the fixed totals.

The rule can now be made that the significance of a table is to be determined by the sum of the probabilities of the table and of all other tables no more probable.

The condition that the probabilities shall not vary from table to table, and the rule just stated, will give a test of significance.

Statistical Flowers Caught in Amber

Paul A. Samuelson

Since I remember well the war-time MIT seminars in statistics now being reproduced in abstract form, I am happy to accept the editors' invitation to reminisce about those times.

Chance alone turned up these Abstracts in the University of Chicago libraries. Although it was my secretary (and Harold Freeman's), Eleanor Prescott Clemence, who typed up these mathematical abstracts, all of us had forgotten they were ever compiled. With probability not minute, Harold Freeman would have sent a copy of them to our friend W. Allen Wallis, who with certainty approaching unity throws away nothing. (The initials W. A. W. on the manuscript Stephen Stigler stumbled upon in the Chicago archives are in the unmistakable schoolboy hand of the Honorable W. Allen Wallis.)

Paul A. Samuelson is Institute Professor Emeritus, Massachusetts Institute of Technology, E52-383C, Cambridge, Massachusetts 02139.

Actually, with faculty blessings, this seminar series was conceived and executed by two graduate students: Lawrence Klein, who was to become MIT's first Ph.D. in Economics and our first home-grown Nobel Laureate; and Joseph Ullman, then studying economics but in the course of the war's windup in Europe later to be enticed into a career in mathematics by Gabor Szegö. Laurie and Joe both as introducers of the speakers; Harold Freeman and I would both cringe and delight in the unpredictable algebraic felicities of their unrehearsed introductions. (Sample: when the illustrious Richard von Mises was to be presented, his many fames as a pioneer had not run ahead of him; so our student impresario left it at, "Although I don't know why, our speaker is supposed to be a very famous scholar.")

It is amazing that, in this epoch after Pearl Harbor, when faculty was dispersing to various war-time labs and graduate student bodies were shrinking to a small core of transients and women, two active students could still attract without stipends so brilliant a group of speakers. Most were