J. NEYMAN

ON THE OCCASION OF HIS 80TH BIRTHDAY

By L. LeCam and E. L. Lehmann

On April 16, 1974 Jerzy Neyman celebrated his 80th birthday. This issue of the *Annals of Statistics* is dedicated to him on this occasion to honor both the scientist, who continues to make important contributions at an astonishing rate, and his earlier work, the impact of which has been so extraordinary that it has completely revolutionized the field of statistics.

At present, Neyman is Director of the Statistical Laboratory at the University of California, Berkeley which he founded in 1938, and is Professor (retired but recalled to active duty) in the Department of Statistics which grew out of this Laboratory in the 1950's. He is supervising several Ph.D. students, has just completed editing a volume of essays on Copernican Revolutions for the National Academy of Science, and is continuing his work on cosmology, weather modification, $C(\alpha)$ tests, and other problems.

Neyman's publications span a period of fifty years.[1] His early work has become so thoroughly part of the common statistical consciousness that it is now only rarely referenced and is no longer conceived as an individual contribution. At the present occasion it may therefore be appropriate to sketch briefly the formative influence which this work has exerted on our discipline.

1. The theory of hypothesis testing developed in collaboration by Neyman and E. S. Pearson[2] in the years 1928–1938 ushered in the subject of mathematical statistics as we know it today. The first of their joint papers [1] brings the fundamental ideas that the choice of a test requires consideration of the *alternatives* to the hypothesis being tested, and that there are *two kinds of error*, false acceptance and false rejection, both of which must be taken into account. Fisher's likelihood ratio is then proposed as an intuitively appealing solution to the testing problem. The rest of the paper and those following in the next few years are devoted to working out the likelihood ratio tests for a number of important examples.

The second decisive step came in the 1933 paper [3], in which the authors are no longer satisfied with an intuitive solution to their problem but take on the task of determining the test which at the given significance level will maximize the power against a given alternative. (Actually, the terms "power," "most powerful," etc. are only introduced in the next paper [4].) This problem is solved completely for the case of a simple hypothesis by the celebrated Neyman-Pearson Fundamental Lemma, which states that for testing a simple hypothesis against a simple alternative the

[1] A bibliography, complete at that time and comprising 156 items, can be found in the volume *A Selection of Early Statistical Papers of J. Neyman*, University of California Press, 1967.

[2] The ten papers comprising this work are reprinted in the volume *Joint Statistical Papers of J. Neyman and E. S. Pearson*, University of California Press, 1966.

solution is indeed given by the likelihood ratio test with the specified significance level.

In the case of composite hypotheses, the restriction to *similar* tests is introduced (i.e., tests whose probability of rejection is the same for all members of the hypothesis). The solution of the problem is then obtained for families of distributions satisfying certain differential equations. (The solutions to these equations constitute in fact what is known today as an exponential family.)

In both the simple and composite cases, it turned out that in some of the most important examples there exists a uniformly most powerful (similar) test, i.e., one that maximizes the power simultaneously against all alternatives of interest. The problem of what to do when no such uniformly best test exists is taken up in [11]. A solution is obtained by imposing on the tests the additional restriction of unbiasedness (i.e., to tests whose power is greater than or equal to the significance level for all alternatives of interest). For one-parameter families, a locally most powerful unbiased test (type A) is derived, and for exponential families this is shown to be UMP unbiased (type A_1). The first of these results is extended to families with nuisance parameters in [9]. The corresponding extension of the second result which would have completed the program was not carried out by the authors; it was furnished later by Scheffé.

The impact of this work has been enormous. It is, for example, hard to imagine hypothesis testing today without the concept of power, which provides the basis both for the determination of sample size and for any comparisons among competing tests. And the optimum properties of the classical normal-theory tests are not only aesthetically pleasing but serve as benchmarks against which the performance of simpler or more robust tests can be gauged. However, the influence of the work goes far beyond its implications for hypothesis testing. By deriving tests as the solutions of clearly defined optimum problems, Neyman and Pearson established a pattern for Wald's general decision theory and for the whole field of mathematical statistics as it has developed since then.

2. One of Neyman's most important creations is the theory of confidence sets. Estimation by confidence sets is today considered one of the classical methods of statistical inference (together with point estimation and hypothesis testing) and is set forth in every textbook of statistics. Although, as with every important idea, there are some forerunners, Neyman was the first in his fundamental paper [12] to give a general formulation and stress its frequency interpretation.

A key feature of the theory developed by Neyman is the relation between confidence sets and tests. Suppose that for each given value θ_0 of the parameter θ being estimated, $A(\theta_0)$ denotes the acceptance region of the hypothesis $\theta = \theta_0$ at significance level α. Then the parameter sets $S(x)$ defined for each sample point x by the relation

$$\theta \in S(x) \Leftrightarrow x \in A(\theta)$$

constitute confidence sets for θ with confidence coefficient $1 - \alpha$.

This relationship makes it possible for Neyman to transfer the concepts and results of the recently developed theory of hypothesis testing to provide an analogous theory of estimation by confidence sets. In particular, the tightness of a confidence statement (which plays a role similar to that of the power of a test) in this theory is measured by the probability of the confidence set $S(X)$ covering values of θ other than the true value.

A brief statement of the idea of confidence intervals is given already in [5], applications are made in [10], and the impossibility of exact (non-randomized) confidence intervals for the binomial case is proved in [8]. A full exposition of the theory is also given in [15], where uniformly shortest unbiased (type A_1) confidence intervals are obtained for a family of distributions which is essentially an exponential family. The relationship between confidence intervals and Fisher's fiducial theory is investigated in [17].

3. While optimum tests or estimates exist and can be obtained explicitly for many important problems, there are also many situations for which this is not the case. In the late thirties and early forties the general purpose estimation procedure for such problems was that of maximum likelihood. The resulting estimates are approximately optimal in large samples but computationally they were frequently quite intractable.

A remedy was provided by J. Neyman, to whom we owe two broad classes of flexible and widely applicable procedures which are still justifiable, even though only asymptotically, from the point of view of the optimal theory. One class of procedures, intended for both estimation and testing purposes is given by Neyman's theory of Best Asymptotically Normal (BAN) estimates. The other class, intended for testing purposes in the presence of nuisance parameters, is the class of $C(\alpha)$ tests.

In a classic paper [18] written before 1945, but published in 1949, Neyman shows that through minimization of appropriately chosen expressions one can obtain classes of estimates which are all asymptotically equivalent and all provide the most concentrated limiting Gaussian distributions. For these, not only the claims of asymptotic optimality could be proved, but also the flexibility afforded by the choice of expressions to be minimized could be used to reduce necessary computations drastically.

Neyman proceeds then to show that these estimates also provide tests procedures which are asymptotically optimal, and all asymptotically equivalent. The study of the large sample performance of these tests is made through consideration of "nearby alternatives" which approach the hypothesis at the rate $1/\sqrt{n}$. (This technique which has invaded most of present asymptotic work had been introduced earlier by Neyman in an ingenious paper [13] on the "smooth" test for goodness of fit.) The theory of BAN estimation has now acquired the same status as that of least squares estimation, and can be regarded as the asymptotic equivalent of this last method.

The theory of $C(\alpha)$ tests derives its importance from the fact that construction of optimal similar tests in the presence of nuisance parameters is often difficult. It is then tempting to solve the simpler problem obtained by replacing the values of the nuisance parameters by estimated values. Unfortunately this does not work. What Neyman does is to show that one can make it work asymptotically by appropriate modification of the test statistics [23]. The optimal tests so constructed are easily obtainable from logarithmic derivatives of the densities. They can be applied to a variety of complex problems which can hardly be tackled otherwise and constitute a delightfully growing chapter of asymptotic theory [36].

4. Throughout his scientific career Neyman has been concerned with statistical problems which arise in various fields of human endeavor. More than half of his published papers are devoted to direct study of specific questions in a variety of domains which can be roughly classified as Agriculture, Biometry, and Health, Astronomy and Meteorology. Some of the early papers, published in Polish, are unfortunately not widely available. The problems they describe were the origin of the soul searching which eventually bore fruit in the form of the theory of testing hypotheses and interval estimation. In the early thirties, Neyman's position as Head of the Nencki Biometric Institute provided him with a multitude of problems which can hardly be described here in recognizable form. Among those which led to publications one could mention the problems of counting the number of viruses or bacteria needed to cause disease [2], the accuracy of the dilution method [10], sickness due to industrial exposure [6], health insurance [7], etc.

The postwar years saw Neyman involved in the design of mass screening for tuberculosis and in the evaluation of treatments against cancer. The fact that each individual is exposed to several causes of death, whose relative contributions to mortality must be disentangled, led Neyman to elaborate what he calls the theory of competing risks together with models of relapse and recovery [19]. Another series of papers deals with accident proneness and the possibility of distinguishing between variability of proneness and contagion phenomena [20, 21]. One can also classify in the biometric domain several papers on population dynamics [25, 26]. Some of these were prompted by the need for explanation of phenomena noticed by T. Park in his populations of flour beetles: two different species (*T. castaneum* and *T. confusum*) seem to survive indefinitely if raised separately. On the contrary, when raised together, one of the two species disappears rather rapidly, the probability of survival of either being a function of the environment provided. Still in the domain of health, and in addition to various studies of models of epidemics [30], Neyman has been involved for many years in research on the mechanism by which neoplastic cells are produced [32]. These studies, and an experiment on the effect of varying doses of urethane in mice, are still continuing [38].

One of Neyman's deepest involvements in a substantive field is a collaboration of more than twenty years with E. L. Scott and many astronomers on problems arising from the clustered appearance of photographs of extragalactic bodies. In addition to statistical approaches to general cosmology [29] and to the study of galactic evolution [34], this collaboration has resulted in an extensive theory of the process of clustering, with the description of clusters of clusters interpenetrating each other [24, 28]. This theory of clustering is applicable to other domains, besides astronomy. It can be applied, for instance, to the study of epidemics [34]. Even before the problem arose in astronomy, Neyman had encountered clustering in the description of abundance of larvae in the field, leading him to introduce the class of distributions called "contagious distributions" [16].

Another domain in which Neyman has been working for more than twenty years is the frustrating and complex one of artificial stimulation of precipitation. Methods of evaluation of the effect of cloud seeding proposed by commercial operators could not be considered scientifically conclusive. After a few years and a few experiments, it seemed that nothing remarkable could be asserted [27]. However, a review of the Swiss hail suppressing experiments indicated that cloud seeding may well be responsible for surprisingly large effects which could be either negative or positive [31]. Detailed examination of other randomized experiments now leave little choice but to conclude that this is indeed the case, with negative effects often extending downwind for hundreds of miles [33]. A recent summary is given in [37].

If the indications provided by this analysis are at all correct, they should lead one to proceed most cautiously in attempts to weather modification. There are other domains where even more caution appears necessary. One of them involves the effects of pollution on health. A volume of the Sixth Berkeley Symposium devoted to this subject indicates how little is known and how necessary is the large scale effort recommended by Neyman in the Epilogue to the volume [35].

5. A central idea of Neyman's approach to mathematical statistics (which was then new but to which we have become so accustomed that it seems trite even to mention it) is the representation of the phenomenon under investigation by means of a mathematical model, which is both simple enough to be tractable and general enough to permit the development of a general theory. This idea constitutes the unifying theme of his *Lectures and Conferences on Mathematical Statistics and Probability Theory* which was first published in 1938 and reissued in a revised and enlarged edition in 1952. In these lectures he sketches his theories of hypothesis testing and confidence estimation against the historical background. He also presents his important contributions to the theory of survey sampling and provides a discussion of spurious correlations, one of his favorite lecture topics. The book provides an easily accessible introduction to his way of thinking as well as to some of his most important concepts and results.

PUBLICATIONS OF J. NEYMAN

[1] 1928 On the use and interpretation of certain test criteria for purposes of statistical inference (with E. S. Pearson). *Biometrika* **20-A** Part I 175–240 and Part II 263–294.

[2] 1931 Counting virulent bacteria and particules of virus (with K. Iwaszkiewicz). *Acta Biologiae Exp.* **6** 101–142.

[3] 1933 On the problem of the most efficient tests of statistical hypotheses (with E. S. Pearson). *Philos. Trans. Roy. Soc. London Ser. A* **231** 289–337.

[4] 1933 The testing of statistical hypotheses in relation to probabilities *a priori* (with E. S. Pearson). *Proc. Cambridge Philos. Soc.* **29** 492–510.

[5] 1934 On the two different aspects of the representative method. *J. Roy. Statist. Soc.* **97** 558–625. (Spanish version of this paper appeared in *Estadística* **17** 587–651, 1959.)

[6] 1934 *Preliminary Report on the Investigation into the Sickness Experience of Workers in Certain Industries* (with K. Iwaszkiewicz; Polish, English summary). Pamphlet of 55 pages published by the Polish Institute for Social Problems, Warsaw.

[7] 1935 On certain details of English national health insurance statistics in connection with the system of management. *Polish Rev. Social Ins.* 1–4.

[8] 1935 On the problem of confidence intervals. *Ann. Math. Statist.* **6** 111–116.

[9] 1935 Sur la vérification des hypothèses statistiques composées. *Bull. Soc. Math. France* **63** 246–266.

[10] 1935 Statistical studies in questions of bacteriology. Part I. The accuracy of the "dilution method" (with T. Matuszewski and J. Supińska). *J. Roy. Statist. Soc. Suppl.* **2** 63–82.

[11] 1936 Contributions to the theory of testing statistical hypotheses (I) Unbiased critical regions of Type A and Type A_1 (with E. S. Pearson). *Statist. Res. Mem.* **1** 1–37.

[12] 1937 Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. Roy. Soc. London Ser. A* No. 767 **236** 333–380.

[13] 1937 "Smooth" test for goodness of fit. *Skand. Aktuarietidskr.* **20** 149–199.

[14] 1938 Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.* **33** 101–116.

[15] 1938 L'estimation statistique traitée comme un problème classique de probabilité. *Actualités Sci. Indust.* No. 739 25–57. (Russian version of this paper appeared in *Uspehi Mat. Nauk* **10** 207–229, 1944.)

[16] 1939 On a new class of contagious distributions, applicable in entomology and bacteriology. *Ann. Math. Statist.* **10** 35–57.

[17] 1941 Fiducial argument and the theory of confidence intervals. *Biometrika* **32** 128–150.

[18] 1949 Contribution to the theory of the chi-square test. *Proc. Berkeley Symp. Math. Statist. Prob.* (1945) 239–273. Univ. of California Press.

[19] 1951 A simple stochastic model of recovery, relapse, death and loss of patients (with E. Fix). *Human Biology* **23** 205–241.

[20] 1952 Contribution to the theory of accident proneness, I. An optimistic model of correlation between light and severe accidents (with Grace E. Bates). *Univ. of Calif. Publ. Statist.* **1** 215–254.

[21] 1952 Contribution to the theory of accident proneness, II. True or false contagion (with Grace E. Bates). *Univ. California Publ. Statist.* **1** 255–276.

[22] 1952 *Lectures and Conferences on Mathematical Statistics and Probability Theory*. Graduate School, U. S. Dept. of Agriculture, Washington, D. C.

[23] 1954 Sur une famille de tests asymptotiques des hypothèses statistiques composées. *Trabajos Estadíst.* **5** 161–168.

[24] 1955 Sur la théorie probabiliste des amas de galaxies et la vérification de l'hypothèse de l'expansion de l'univers. *Ann. Inst. H. Poincaré Sect. B* **14** 201–244.

[25] 1956 Struggle for existence. The *Tribolium* model: Biological and Statistical aspects (with Thomas Park and E. L. Scott). *Proc. Third Berkeley Symp. Math. Statist. Prob.* **4** 41–79. Univ. of California Press.

[26] 1959 Stochastic models of population dynamics (with E. L. Scott). *Science* **130** 303–308.

[27] 1960 Chapter V—Santa Barbara randomized cloud seeding project—Evaluation of seeding operations, Santa Barbara and Ventura Counties in 1957, 1958 and 1959 (with E. L. Scott and M. Vasilevskis). *Interim Report on the Santa Barbara Weather Modification Project* Department of Water Resources, Sacramento, California, pp. VI–V56.

[28] 1962 Alternative stochastic models of the spatial distribution of galaxies. *Problems of Extragalactic Research* (I.A.U. Symposium 15). 294–314. Macmillan, New York.

[29] 1963 Stochastic approach to cosmology. *Proc. of the Conference on Mathematical Models in Physical Sciences*. 135–174. Prentice Hall, New York.

[30] 1964 A stochastic model of epidemics. *Stochastic Models in Medicine and Biology* (with E. L. Scott). 45–83. Univ. of Wisconsin Press.

[31] 1967 Statistical aspects of the problem of carcinogenesis (with E. L. Scott). *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4** Univ. of California Press.

[32] 1967 Some outstanding problems relating to rain modification (with E. L. Scott). *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **5** Univ. of California Press.

[33] 1971 Further studies of the Whitetop cloud seeding experiment (with J. L. Lovasich, E. L. Scott and M. A. Wells). *Proc. Nat. Acad. Sci. U. S. A.* **68** 147–151.

[34] 1972 Processes of clustering and applications (with E. L. Scott). *Proceedings, Symposium on Stochastic Point Processes*, (P. A. W. Lewis, ed.). 646–681. Wiley, New York.

[35] 1972 Epilogue of the health-pollution conference. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **6**; *Effects of Pollution on Health*, (L. LeCam, J. Neyman and E. L. Scott, eds.). Univ. of California Press, Berkeley and Los Angeles (1972) 575–589; and *Bull. Atomic Sci.* **19** (1973) 25–34.

[36] 1972 $C(\alpha)$ tests and their use. Proceedings of the Conference of the Advanced Institute of Statistical Ecology in the United States, Pennsylvania State Univ., July 1972. (Unpublished.)

[37] 1973 Some current problems of rain stimulation research (with E. L. Scott). *Proceedings of the Symposium on Uncertainties in Hydrologic and Water Resources Systems*, (C. C. Kisiel, Chairman). Univ. of Arizona. 1167–1244.

[38] 1973 A view of biometry: an interdisciplinary domain concerned with chance mechanisms operating in living organisms. Illustration: urethan carcinogenesis. To be published in *Proc. Conf. Biometrics*, Florida State Univ. July, 1973.