References

[A-W] H. Amann and S. Weiss, *On the uniqueness of the topological degree*, Math. Z. **130** (1973), 39–54.

[B-K] G. D. Birkhoff and O. D. Kellogg, *Invariant points in function spaces*, Trans. Amer. Math. Soc. **23** (1922), 96–115.

[E] I. Ekeland, *Nonconvex minimization problems*, Bull. Amer. Math. Soc. (N.S.) **1** (1979), 443–474.

[P] W. V. Petryshyn, *On the approximation solvability of equations involving A-proper and pseudo-A-proper mappings*, Bull. Amer. Math. Soc. **81** (1975), 223–315.

P. M. Fitzpatrick
University of Maryland

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 20, Number 2, April 1989

0273-0979/89 $1.00 + $.25 per page

Asymptotic methods in statistical decision theory, by Lucien Le Cam. Springer Series in Statistics, Springer-Verlag, New York, Berlin, Heidelberg, 1986, xxvi + 742 pp., $49.95. ISBN 0-387-96307-3

It is rather odd that in statistics no "general" theory has yet been formulated that would investigate from a single point of view the properties of estimates, tests and other inference procedures and would be accepted by the majority of specialists. A fundamental concept such as optimality is usually defined *ad hoc* in each special case, depending on the nature of the problem. Therefore, the number of concepts of optimality in statistics grows along with the number of problems treated and models studied. The work of A. Wald, D. Blackwell and others in abstract statistical decision theory has shown how successfully the ideas and methods of the theory of optimization and convex analysis work in statistics. This has led to some systematization of the various concepts of optimality. The development of the asymptotic theory of statistical inference has shown that under mild regularity assumptions various concepts of optimality often lead to the same or almost the same solutions. These ideas have been especially well studied in the more traditional schemes of statistical inference, in which for n independent identically distributed (i.i.d.) observations $X^n = (X_1, \dots, X_n)$ representing a sample of size n from the distribution F_θ, one wishes to estimate the value of the parameter θ or derive a test for verifying some hypothesis about this parameter. The similarity, in such a situation, of these different approaches to the concept of asymptotic optimality is explained by the fact that they all depend on using an appropriate form of the scaled likelihood ratio

$$\frac{dP^n_{\theta+h/\sqrt{n}}}{dP^n_\theta}(X^n) \equiv Z^n_h \equiv Z^n_h(\theta; X^n),$$

where P^n_θ is the distribution of X^n derived from F_θ. Let

$$Z_h = \exp(L_\theta(h) - \tfrac{1}{2}(J_\theta h, h))$$

denote the likelihood ratio for a Gaussian experiment with unknown shift parameter h. Convergence of the finite-dimensional distributions of the local likelihood ratio process $Z^n_\cdot$ to the finite-dimensional distributions of the process $Z_\cdot$, i.e.

$$Z^n_\cdot \xrightarrow{D_f} Z_\cdot,$$

has received the name of local asymptotic normality (LAN). The LAN property has stimulated the development of a rather powerful "general" theory of asymptotic optimal statistical solutions based on the independence of repeatable experiments when the sample size $n \to \infty$. The fundamental consequence of LAN appears in the inequalities of Hájek-Le Cam, in Hájek's convolution theorem, and in other theorems which allow us to describe the asymptotic properties of the minimax risk of procedures in terms of the analogous risk for the limiting Gaussian experiment. These results play a notable role not only in parametric problems in statistics, but also in many nonparametric and semiparametric problems as well. (We recall, for example, that in i.i.d. schemes of observations the nonparametric approach assumes that θ is infinite-dimensional while the semiparametric approach usually is characterized by a representation $\theta = (\theta_1, \theta_2)$, where the finite-dimensional component θ_1 is the estimated parameter and the infinite-dimensional component θ_2 is a nuisance parameter.)

The achievements in the asymptotic theory of these problems are connected with the names of C. Stein, J. Hájek, L. Le Cam, P. Bickel, R. Beran, I. A. Ibragimov, R. Z. Has′minskii, D. M. Chibisov, P. W. Millar, J. Wellner, J. Pfanzagl, B. Ya. Levit, and other authors. However, one should note that asymptotic normality is not the only possibility for the limiting behavior of a likelihood ratio. Even in an i.i.d. scheme of observations, when some of the usual conditions of regularity are violated, one can use a normalization (δ_n^{-1}) that is different from $\sqrt{n}$, to establish that as $n \to \infty$ the finite-dimensional distributions of the processes

$$Z^n_h(\theta) = \frac{dP_{\theta+\delta_n h}}{dP^n_\theta}(X^n)$$

converge weakly to the corresponding finite-dimensional distribution of the likelihood ratios

$$Z_h(\theta) = \frac{dP_{\theta+h}}{dP_\theta},$$

for a non-Gaussian infinitely divisible experiment. On the other hand, the LAN property with suitable normalizations can be established in a variety of situations where the observations are not necessarily identically distributed or where even dependent observations are possible; such situations can occur in problems involving statistics of random sequences and processes. In particular, in appropriate situations the reasoning of the asymptotic theory in the "pure" form, not burdened with excessive assumptions of regularity, clarifies the concept of asymptotic normality of experiments and often also suggests the form of the optimal solution.

The book of Professor Le Cam is devoted to an exposition of asymptotic statistical decision theory in an even more general form. Many of the results of the book are already well known, because of numerous applications, at first by the author himself, but also by other well-known statisticians who have referred to other books, papers and preprints, as well as an unpublished variant of the book currently under review. Apparently considering the style and contents of the book as a whole to be more or less known to a sufficient number of specialists, the author introduces general concepts in the first paragraphs of the book. The author develops the foundations of the modern theory of statistical decisions—experiments, the decision procedures, the risk function, the distance between experiments, weak convergence of experiments, etc. The endeavor for generality distinguishes Professor Le Cam himself as well as the book under review. But in connection with the book, this property at times gives rise to some criticism. For example the reader may be somewhat disheartened that the well-known notion of statistical decision rule is defined not as a Markov kernel, $T(x, A)$ (which has very clear meaning as a conditional probability that a decision takes values in the set A when the observation is x), or even as a bilinear form on the cartesian product of the space $M(x, A)$ of measures of bounded variation and the space of functions Γ on the initial set D of possible solutions. The author introduces the statistical decision as an operator $T: \Gamma^* \to L(\mathscr{E})$ from the dual Γ^* of the uniform lattice Γ to the abstract L-space $L(\mathscr{E})$ generated by the experiment $\mathscr{E}$, although a few pages later, in §I.6 (pp. 11–15) it turns out that in principle everything can be reduced to the investigation of Markov kernels. In our view, it would have been preferable to begin the exposition by using the more usual concept and go from the special to the general. One can make analogous notes to other chapters of the book. At the same time, the systematic use of many general concepts such as the conical measure induced by an experiment, Choquet lattices, etc., allows one to describe the risks of procedures and study LAN conditions and Δ-convergence of experiments in the most general form. The concept of Δ-convergence of experiments, which also arose in the works of Le Cam, means something more than LAN, namely that the minimax risks of a sequence of problems converge to the corresponding risks for the limit experiment. Therefore, one can extend the theorems on lower bounds of asymptotic risk of statistical procedures to practically all conceivable statistical problems. This reduces theorems on asymptotic minimax to a general principle, the "law of nature" of statistical solutions.

In the first ten chapters of the book, this analytical apparatus is developed. Along with the concepts from statistical theory recalled above, it is related to criteria of sufficiency and asymptotic sufficiency generalizing the Halmos-Savage theorem, to the convergence of distributions of procedures and approximation estimates in probability for tests, and to the Ionescu-Tulcea theorems on lifting, which give the existence of invariant or equivariant rules. In addition, the typical limiting experiments, Gaussian and Poisson, are described here, and the characterization of infinitely divisible experiments is introduced, along with an exposition of a related modification of the central limit theorem.

The following four chapters (10–13) are devoted to the general problems of asymptotic normality. It is clear that appropriate asymptotic approximations of statistical problems by problems involving shifts of the Gaussian distribution simplifies substantially the study of asymptotic optimality. Roughly speaking, we know that maximum likelihood estimates and likelihood ratio tests are "good" for the standard Gaussian experiment. Therefore, if under a suitable normalization the experiments $\mathscr{E}^n$ are approximated by a standard Gaussian experiment $\mathscr{E}$ in the same sense as for i.i.d. observation schemes, then one can establish the usual limit properties of the maximum likelihood estimate and other "standard" procedures of statistical inference. General results are exposed such as theorems on asymptotic admissibility and minimax solutions for experiments converging to a standard Gaussian one. Thus not only in the i.i.d. scheme of observations, but wherever it is possible to establish, in an appropriate form, the convergence of normalized likelihood ratios of those of Gaussian experiments, one succeeds in obtaining lower bounds for the asymptotic risk. For example such bounds are used when we try to consider a shrinking sequence of neighborhoods for some given point in the parameter set. For traditional schemes of i.i.d. observations, Le Cam had proposed a general method of constructing asymptotically efficient procedures of estimation. Later P. W. Millar proposed using such a construction in nonparametric estimation. Chapter 11, where the global theory of asymptotic normality is developed, is devoted to attempts at a further generalization of a similar construction to the general problem. However the theoretical part prevails over the "practical" achievements here. At the same time, the global theory of asymptotic normality is an important method, oriented not merely to the description of lower bounds of risks, but rather permitting one to realize in a general form the process of constructing optimal solutions. For this reason the "standard" method of statistical inference (involving the usual maximum likelihood estimates) has not yet been developed for more general schemes of nonparametric statistics. In any case their use is constrained by essential analytic limitations. The abstract constructions exposed in Chapter 11 generalizing earlier work of Le Cam himself to nontraditional schemes of observation, deserve consideration. The global theory studies not so much the approximation by one Gaussian experiment as the local approximation of a sequence of experiments $\mathscr{E}^n$ by a sequence of Gaussian experiments $\mathscr{G}^n$. Just such a situation arises in the case of i.i.d. observations when as a preliminary step, one constructs a $\sqrt{n}$-consistent estimate $\overline{\theta}^n$ (or some discretization of it), and then, assuming one is in a neighborhood of $\overline{\theta}^n$, one uses an appropriate quadratic approximation of the logarithm of the likelihood ratio, which allows one to apply stochastic approximation methods or the Newton-Raphson method. Exposing this asymptotic theory, the author points out the important role that martingale methods can play in deriving the asymptotic normality of experiments. This reasoning becomes even more important for the investigation of inference problems involving stochastic processes, especially processes with discrete time. Then, as is shown in several works on limit theorems for semimartingales, one may obtain criteria for asymptotic normality of experiments in "predictable terms," where predictability is understood in the sense of the "general theory of stochastic processes."

In Chapter 12 the limit properties of posterior distributions and Bayesian procedures are described and in Chapter 13 some results relating to sequence of experiments. It is possible that some of the general principles formulated here will be more fully developed through further research on concrete statistical procedures.

The remaining part of the book is connected with the more traditional schemes of statistical theory. The author exposes such topics as approximation of exponential families of distributions and the related problems of asymptotic sufficiency (Chapter 14) and also several topics concerning sums of independent random variables (Chapter 15). The analysis of the asymptotic behavior of various procedures, related with generalized regression models, is based on results developed in these chapters. The independence of observations allows one to obtain some specific limit theorems; in particular a construction of Poisson approximation, or "Poissonization", is described which is useful in applications. The two final chapters (16 and 17) are devoted to the direct application of results on independent observations in statistics. In particular, the asymptotic behavior of estimation procedures is studied. The asymptotic normality of the procedures is established, formulas are deduced for limiting risks, and methods for constructing optimal estimates and tests are described (Chapter 16). In Chapter 17 the corresponding conclusions are obtained for traditional schemes of i.i.d. observations. Here, general results are established about the existence of $\sqrt{n}$-consistent estimates. They are obtained in terms of the Hellinger distance for the initial experiment. Asymptotically efficient estimates are often constructed by the one-step approximation. Such a general method was introduced by Le Cam. Thus these results together with contents of Chapter 17 are a foundation of a universal methodology for statistical inference in the traditional i.i.d. case.

In the appendices, all the facts and constructions from general topology, the theory of sets and functions, measure theory, and functional analysis, which are employed in this exposition of general statistical decision theory are collected.

The book of Professor Le Cam presents a detailed exposition of the achievements of abstract statistical decision theory. Many, if not all, of the chapters of this book are directly or indirectly related to the work of Professor Le Cam himself, and the style of exposition carries the mark of the attractive personality of the author. It is possible that another author of a similar book would have put its stress elsewhere. In this connection, we would like to contrast the book under review with other works on "general" theory. Thus the chapters of the book which contain the general results on asymptotic sufficiency and on asymptotically exponential families of distributions are a good complement to the well-known book of N. N. Chentsov, *Statistical decision rules and optimal inferences* ("Nauka", Moscow, 1972), where a differential-geometric approach prevails and the equivalence of experiments is treated in the language of geometry, constructed with the help of morphisms or Markov kernels. The material of Chapters 14–17 can be considered as a deeper exposition of the facts whose development is the subject of the book of I. A. Ibragimov and R. Z. Has′minskii, *Statistical estimation. Asymptotic theory* (Springer-Verlag, 1981). The theorems on asymptotic minimax were the special theme of the work of P. W. Millar, *The minimax principle in asymptotic statistical theory* (Springer Lecture Notes in Mathematics, No. 976). The concretization proposed by P. W. Millar of the general scheme of Le Cam permitted an elegant exposition of the results of the theory of asymptotic optimality of nonparametric estimation of R. Beran, B. Ya. Levit, and others. In particular, it allows one to obtain a minimax property for the empirical distribution function under mild assumptions on the loss function. A book appeared quite recently which is especially close to the contents of the book under review, *Mathematical theory of statistics* by Helmut Strasser (Walter de Gruyter, Berlin, New York, 1985), in which examples of the problems of hypothesis testing and estimation procede the exposition of the abstract concepts. This latter book is worth recommending for a preliminary acquaintance to those who would wish to study and apply the results presented in Le Cam's book.

On the whole, despite certain difficulties in reading, related primarily to the predominant role of the abstract exposition, the book of Professor Le Cam represents an important step in the formulation of a General Theory of Statistical Inference. The limit theorems of mathematical statistics presented here are based on minimal assumptions and no doubt will find wide application in diverse problems of statistics.

(Review translated from the Russian by James R. King)

A. N. SHIRYAYEV AND YU. A. KOSHEVNIK
V. A. STEKLOV MATHEMATICAL INSTITUTE