Springer Series in Statistics

Lucien Le Cam

Asymptotic Methods in Statistical Decision Theory

统计决策理论中的渐近方法

Springer

世界图书出版公司
www.wpcbj.com.cn

Lucien Le Cam

Asymptotic Methods in Statistical Decision Theory

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo

图书在版编目（CIP）数据

统计决策理论中的渐近方法 = Asymptotic Methods in Statistical Decision Theory : 英文 /（美）L. 勒卡姆（L. Le Cam）著 .—影印本 .—北京：世界图书出版公司北京公司，2016.11
ISBN 978-7-5192-2079-2

Ⅰ．①统…　Ⅱ．①勒…　Ⅲ．①统计决策理论—研究—英文
Ⅳ．① O212.5

中国版本图书馆 CIP 数据核字（2016）第 272592 号

著　　者：Lucien Le Cam
责任编辑：刘　慧　高　蓉
装帧设计：任志远

出版发行：世界图书出版公司北京公司
地　　址：北京市东城区朝内大街 137 号
邮　　编：100010
电　　话：010-64038355（发行）　64015580（客服）　64033507（总编室）
网　　址：http://www.wpcbj.com.cn
邮　　箱：wpcbjst@vip.163.com
销　　售：新华书店
印　　刷：三河市国英印务有限公司
开　　本：711mm × 1245mm　1/24
印　　张：32
字　　数：605 千
版　　次：2017 年 1 月第 1 版　　2017 年 1 月第 1 次印刷
版权登记：01-2016-6784
定　　价：115.00 元

Springer Series in Statistics

Advisors:
D. Brillinger, S. Fienberg, J. Gani,
J. Hartigan, K. Krickeberg

Springer Series in Statistics

D.F. Andrews and A.M. Herzberg, Data: A Collection of Problems from Many Fields for the Student and Research Worker. xx, 442 pages, 1985.

F.J. Anscombe, Computing in Statistical Science through APL. xvi, 426 pages, 1981.

J.O. Berger, Statistical Decision Theory and Bayesian Analysis, 2nd edition, xiv, 425 pages, 1985.

P. Brémaud, Point Processes and Queues: Martingale Dynamics. xviii, 354 pages, 1981.

K. Dzhaparidze, Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. xii, 300 pages, 1985.

R.H. Farrell, Multivariate Calculation. xvi, 367 pages, 1985.

L.A. Goodman and W.H. Kruskal, Measures of Association for Cross Classifications. x, 146 pages, 1979.

J.A. Hartigan, Bayes Theory. xii, 145 pages, 1983.

H. Heyer, Theory of Statistical Experiments. x, 289 pages, 1982.

I.T. Jolliffe, Principal Component Analysis. xiii, 272 pages, 1986.

M. Kres, Statistical Tables for Multivariate Analysis. xxii, 504 pages, 1983.

M.R. Leadbetter, G. Lindgren and H. Rootzén, Extremes and Related Properties of Random Sequences and Processes. xii, 336 pages, 1983.

L. LeCam, Asymptotic Methods in Statistical Decision Theory. xxvi, 742 pages, 1986.

E.B. Manoukian, Modern Concepts and Theorems of Mathematical Statistics. xiv, 156 pages, 1986.

R.G. Miller, Jr., Simultaneous Statistical Inference, 2nd edition. xvi, 299 pages, 1981.

F. Mosteller and D.S. Wallace, Applied Bayesian and Classical Inference: The Case of *The Federalist* Papers. xxxv, 301 pages, 1984.

D. Pollard, Convergence of Stochastic Processes. xiv, 215 pages, 1984.

J.W. Pratt and J.D. Gibbons, Concepts of Nonparametric Theory. xvi, 462 pages, 1981.

L. Sachs, Applied Statistics: A Handbook of Techniques, 2nd edition. xxviii, 706 pages, 1984.

E. Seneta, Non-Negative Matrices and Markov Chains. xv, 279 pages, 1981.

D. Siegmund, Sequential Analysis: Tests and Confidence Intervals. xii, 272 pages, 1985.

V. Vapnik, Estimation of Dependences Based on Empirical Data. xvi, 399 pages, 1982.

K.M. Wolter, Introduction to Variance Estimation. xii, 428 pages, 1985.

A.M. Yaglom, Correlation Theory of Stationary and Related Random Functions. x, 640 pages, 1986.

Lucien Le Cam
Department of Statistics
University of California, Berkeley
Berkeley, CA 94720
U.S.A.

AMS Classification: 62LXX

Library of Congress Cataloging in Publication Data
Le Cam, Lucien M. (Lucien Marie), 1924–
Asymptotic methods in statistical theory.
(Springer series in statistics)
Bibliography: p.
Includes index.
1. Mathematical statistics—Asymptotic theory.
I. Title. II. Series.
QA276.L335 1986 519.5 86-10055

The present version was prepared by Ms. "Bionic Fingers" Ruth Suzuki, whose patience and skill are most extraordinary indeed.

Finally, I wish to thank all the students who suffered through my lectures on the subject.

One of them, Mr. Yu-Lin Chang, took the trouble to read most of the manuscript with considerable care, thus saving me from the embarrassment of very many errors. There are certainly some left, but this is not his fault. Another student, Dr. Yannis Yatracos, compiled the references and the start of an index. Some other students contributed easily identifiable results. If these are mentioned here, I tried to give credit. In some other cases, it is more difficult to give credit because the influence is more spread out. For instance, I owe a special debt to Grace Yang, who in 1966–67 made me realize that even in the locally asymptotically normal case it can be advantageous to treat the problem as if it was what is now called "the random information case." I should have known that, but a glance at the Appendices of my 1960 paper shows that I missed the boat. Grace's suggestion turned out to be most valuable.

Some of the results reported here were, at appropriate times, the subject of research efforts supported by various agencies. We are particularly indebted to the Alfred P. Sloan Foundation, the U.S. National Science Foundation, the U.S. Army Research Office, the Office of Naval Research, and to a lesser but non-negligible extent to the National Institutes of Health.

For a different kind of support, I am most indebted to my wife, Louise, who rarely complained about my idiosyncracies or about the considerable amount of family time I spent on the present subject.

Finally, I must thank the editorial and production departments at Springer-Verlag for their patience, their help, and a superb final product.

Lucien Le Cam

To the Reader

1. Introduction

This book grew out of lectures delivered at the University of California, Berkeley, over many years. The subject is a part of asymptotics in statistics, organized around a few central ideas. The presentation proceeds from the general to the particular since this seemed the best way to emphasize the basic concepts. The reader is expected to have been exposed to statistical thinking and methodology, as expounded for instance in the book by H. Cramér [1946] or the more recent text by P. Bickel and K. Doksum [1977]. Another possibility, closer to the present in spirit, is Ferguson [1967]. Otherwise the reader is expected to possess some mathematical maturity, but not really a great deal of detailed mathematical knowledge. Very few mathematical objects are used; their assumed properties are simple; the results are almost always immediate consequences of the definitions. Some objects, such as vector lattices, may not have been included in the standard background of a student of statistics. For these we have provided a summary of relevant facts in the Appendix.

The basic structures in the whole affair are systems that Blackwell called "experiments" and "transitions" between them. An "experiment" is a mathematical abstraction intended to describe the basic features of an observational process if that process is contemplated in advance of its implementation. Typically, an experiment consists of a set Θ of theories about what may happen in the observational process. Each theory $\theta \in \Theta$ specifies a probabilistic model for the behavior of the things under observation. This is summarized by a probability measure P_θ on a certain σ-field $\mathscr{A}$. The σ-field is independent of θ. It is the collection of events for which the available instruments permit a determination of whether or not the event occurs. For technical reasons we have used a slightly different set-up, but this does not change anything of practical relevance. It was just more convenient to rewrite the definitions in such a way that the desired theorems are always true, instead of imposing restrictive conditions on a more standard system.

We have taken a restricted view of the aims of the statistician who contemplates performing an experiment. His goals are described in the framework of Wald's decision theory, with loss functions, risk functions, and related objects.

The main underlying theme can then be described as follows: The statistician faces a certain experiment $\mathscr{E} = \{P_\theta; \theta \in \Theta\}$. It may be a rather complex affair. He proceeds by searching for another experiment $\mathscr{F} = \{Q_\theta; \theta \in \Theta\}$ that is (1) known and mathematically tractable, and (2) close enough to $\mathscr{E}$ so that techniques selected for $\mathscr{F}$ can be transformed into techniques for $\mathscr{E}$ without dire consequences.

There are very few known and tractable experiments. The one-dimensional Gaussian shift family may qualify. The same cannot be said of Gaussian shift families in higher dimensions. In spite of this, the large amount of space devoted here to Gaussian shift families is mostly a consequence of other facts. They impose themselves on one from many directions, such as the use of Taylor expansions, invariance properties, observational processes where each individual observation gives only a tiny bit of information, etc. Some of these reasons are detailed in Chapters 8, 9, and 10.

We have mentioned "approximating" an experiment $\mathscr{E}$ by another experiment $\mathscr{F}$. To make sense of this, a distance (or more precisely, a pseudo distance) between experiments has been introduced. This could be done in several manners. One possibility was to use only loss functions that are bounded by unity and insist that $\mathscr{E}$ is close to $\mathscr{F}$ if any risk function achievable on $\mathscr{E}$ is uniformly close to a risk function of $\mathscr{F}$, and conversely. Another possibility was to say that $\mathscr{E}$ and $\mathscr{F}$ are close to each other if each of the two can be very closely mimicked by a randomization of the other. It turns out that the two distances so defined are not only related, but exactly equal.

From time to time results are stated as limit theorems obtainable as something called n "tends to infinity." This is especially so in Chapter 7 where the results are just limit theorems. Otherwise we have made a special effort to state the results in such a way that they could eventually be transformed into approximation results. Indeed, limit theorems "as n tends to infinity" are logically devoid of content about what happens at any particular n. All they can do is suggest certain approaches whose performance must then be checked on the case at hand. Unfortunately the approximation bounds we could get were too often too crude and cumbersome to be of any practical use. Thus we have let n tend to infinity, but we would urge the reader to think of the material in approximation terms, especially in subjects such as the ones described in Chapter 11.

The main body of the theory makes no reference to independence of observations or any such structural aspects of the experiments under consideration. It is therefore applicable to many domains involving stochastic processes and other similar affairs. However, a great deal of the argumentation is local, with approximations valid only in small neighborhoods. To piece these together (as in Chapter 11) requires the use of good auxiliary estimates. We have demonstrated the existence of such estimates only in situations that involve independent observations. One could do it in other contexts, for instance for Markov processes. However, recent results of L. Birgé convinced me that it could be premature to describe such situations at this time.

The passage from purely local results to global ones is described in Chapter 11 for the Gaussian approximations. One can do something similar for other exponential families. Some indications are given in Chapter 14, but a full theory is not developed there.

Chapters 16 and 17 treat, in some detail, experiments in which one observes independent or independent identically distributed random variables. This is mostly to illustrate the general theory and show that it encompasses a sizeable portion of standard asymptotics.

Even in the standard independent identically distributed case no attempt has been made to cover the entire field. A compendium of the results available for that case would require many volumes. The reader will find supplementary material in the book by Serfling [1980]; it hardly overlaps with this volume. Further material can be found in the book by I.A. Ibragimov and R.Z. Has'minskii [1979], whose spirit is closer to ours. At one time we had intended to give an account of the results on "second order efficiency" in line with the concepts developed here. This would take much space. For these results the reader is referred to the works of J. Pfanzagl [1980] and of M. Akahira and K. Takeuchi [1981].

2. Summary of Contents and Historical Notes

This is a brief description of the contents of this volume, chapter by chapter, with a few historical remarks. To mention "history" is to imply attributions of priority. We have not intentionally failed to attribute priorities, nor did we intentionally misattribute them. If that has happened by chance, we shall take refuge behind Steve Stigler's "law of eponymy" [1979] which states that "No scientific discovery is named after its original discoverer."

Chapter 1 gives the basic definitions involving experiments, decision spaces, and transitions. The framework is not far removed from that used by A. Wald [1950], except that we remain in the nonsequential case. Also, we elected to work with vector lattices and positive linear operators instead of families of measures and Markov kernels. The main reason for this choice is that it allows one to work essentially as if all the objects used were finite sets. The basic compactness results (Theorem 2, Chapter 1) and the minimax theorem (Theorem 1, Chapter 2) are always true, instead of being encumbered by unappetizing *ad hoc* assumptions.

The necessary elements of the theory of vector lattices can be found in the Appendix. They were drawn from the papers of S. Kakutani [1941a] [1941b], J. Dieudonné [1944], and F. Riesz [1940]. A short account occurs in Bourbaki's *Integration*, Chapter 2. There are now extensive book-length accounts of the subject. Two excellent references are H.H. Schaefer [1974] and D. Fremlin [1974].

An important technical result of Chapter 1 is Theorem 1. It says that every transition can be approximated by very simple Markov kernels. The result is a strengthened version of a theorem in Le Cam [1964]. The strengthening was performed to accommodate a request by W. Moussatat.

For readers who prefer to work with measurable spaces and Markov kernels we have provided a translation in the form of representation theorems. For further information regarding the representation of positive operators by Markov kernels, see Sazonov [1962] and L. Schwartz [1973a] [1973b].

The name "experiment" for a family of probability measures seems to have been coined by D. Blackwell [1951]. The concept itself is much older; it was familiar to Wald [1939]. One could perhaps give part of the credit for its elaboration to Neyman and Pearson who repeatedly, and against R. A. Fisher's blandishments (see Fisher [1955], page 70), emphasized the necessity of considering all reasonable alternatives.

Chapter 2 gives some basic results from the theory of statistical decision functions (Wald [1950]), from the theory of comparison of experiments, and from the theory of approximation of experiments (Le Cam [1964]). Theorem 1 is a form of the minimax theorem akin to that given in Le Cam [1955]. It is followed by the attendant results on completeness of classes of procedures directly related to Bayes procedures. The initial minimax theorem is that of von Neumann [1928]. Many authors, including J. Ville [1938], Wald [1945], Karlin [1950], and Kneser [1952] gave extensions and modifications of the result. There are better minimax theorems due to M. Sion [1958]. See also C. Berge and Ghouila-Houri [1962].

Comparison of experiments was investigated by Bohnenblust, Shapley, and Sherman [1950] following an idea of von Neumann. The statistical wording with the relation to "sufficiency" is due to Blackwell [1951] [1953] and Stein [1951]; see also C. Boll [1955]. The idea of "deficiency" and distances is taken from Le Cam [1964]. The basic Theorem 2 says that several definitions of "deficiency" are all equivalent. It is taken from Le Cam [1964]. The elaboration in terms of conical measures (Choquet [1969]) was influenced by the work of Torgersen [1970]. The chapter also contains some results on the minimal forms of experiments. These are directly related to the idea of sufficiency.

Chapter 3 revolves around distributions of likelihood ratios and their connections with deficiencies or distances between experiments. It uses Blackwell's "canonical measures." Theorem 1 shows that, for finite parameter sets, the distance between experiments is bounded by the dual Lipschitz distance between their respective canonical measures. This has many implications. One of them is a compactness result, given in Theorem 1 and Theorem 2. The compactness result for finite parameter sets occurs in Le Cam [1969].

The general case, for the weak topology, was stated in Le Cam [1972]. That the "proof" given there is wrong was pointed out to me by W. Moussatat. I then gave a more elaborate proof using representations of conical measures along the lines of the last part of Chapter 2. The proof given here was obtained independently of the work of E. Siebert [1979], but, except for matters of notation and terminology, it appears analogous to that of Siebert.

Proposition 1, relating deficiencies and testing deficiencies, is taken from Torgersen [1970]. It implies the characterization of equivalence of experiments by isometries of their linear spans (Proposition 2). This characterization is given in Le Cam [1964] and in Torgersen [1970]. Theorem 3, characterizing boundedly complete experiments as extreme points, is imitated from Torgersen [1977]. Proposition 4 involves direct products of experiments and the fact that making a direct product amounts to convolute conical measures. The convolution operation admits a representation by pointwise products of the associated Hellinger transforms. Theorem 4 is a form of the Blackwell–Sherman–Stein theorem. For other related results, see V. Strassen [1965].

Chapter 4 gives a variety of inequalities. The inequalities between Hellinger distances and L_1-norms occur in C. Kraft [1955]. The relations with the chi-square like distance (called $k(P, Q)$) are easy, but we did not find them in the literature. The distance called $k(P, Q)$ occurs in a paper of Sanghvi [1953]. Proposition 2 and some of the results that follow it are essential in that they give relations between proximity of measures, as computed by L_1-norms, and proximity of likelihood ratios. The inequalities called (*a*) or (*d*) imply what is called Scheffé's theorem (Scheffé [1947]). The inequalities between conditional expectations (Proposition 3 and 4) will be used in connection with studies of "insufficiency" (Chapter 5) and convergence of posterior distributions (Chapter 12). Proposition 5 is about what happens if one uses for a pair (P_0, P_1) tests that were designed to be optimal for a pair (G_0, G_1) with small dual Lipschitz distances $\|P_i - G_i\|_D$.

Chapter 5 concerns sufficiency and approximate sufficiency. For a dominated experiment $\mathscr{E} = \{P_\theta;\ \theta \in \Theta\}$ on a σ-field $\mathscr{B}$, a sub σ-field $\mathscr{A} \subset \mathscr{B}$ is sufficient if and only if the experiment obtained by restricting the P_θ to $\mathscr{A}$ is equivalent to $\mathscr{E}$ itself. This was proved by Bahadur [1955] but had escaped my attention when I wrote Le Cam [1964]. It is *not* a direct consequence of the Blackwell–Sherman–Stein theorem. Here we list eight different properties and show that they are all equivalent to sufficiency. To avoid some technicalities and paradoxes (Burkholder [1961]) we work with positive projections in vector lattices instead of conditional expectation operations. The literature on such projections is large. See, for instance, Neveu [1972] and the references given therein.

If a σ-field $\mathscr{A} \subset \mathscr{B}$ is not sufficient there are several ways of measuring how far it is from being sufficient. One of them is the deficiency of Chapter 2. One can also see how much the P_θ should be modified to make $\mathscr{A}$ sufficient. This gives a different number called "insufficiency" here. Actually there are several variations on that theme and we were not able to show that they are all identical. At any rate they give numbers larger than the deficiencies. Examples are given of this, as well as theorems that allow bounding insufficiencies in terms of deficiencies in special cases. See, for instance, Proposition 5. Since deficiencies or insufficiencies can often be evaluated only on very small parts of the parameter spaces, a machine to put these local evaluations together is needed. Propositions 6 and 7 give some results in that direction. Proposition 8 corrects a wonderfully wrong inequality of Le Cam [1974].

Chapter 6 is about compactness and contiguity. It lists a theorem of Dunford, Pettis, and Grothendieck [1953] [1955]. (The proof is in the Appendix.) It also proves the equivalence of five definitions of contiguity and some useful corollaries of the definitions. Propositions 5 and 6, dealing with joint limiting distributions, are easy. Special cases have appeared in the literature. (See Le Cam [1960], Hall and Loynes [1977] for instance.) The special case of limiting Gaussian distributions is briefly mentioned in Proposition 7. Other results of this type, for infinitely divisible limit distributions, are given in Chapter 16 as well as in Le Cam [1960]. A particular case occurs in Behnen and Neuhaus [1975]. The chapter ends with some results of Lindae [1972]. Theorem 1 of Section 4 is particularly important. It is one of very few links between the weak and strong convergences of experiments.

As explained earlier in this Introduction, we attempted to state results as much as possible so that they could be read as *approximation* theorems instead of *limit* theorems. Chapter 7 is an exception. It deals exclusively with limit theorems. Proposition 1 is about a relation between limit experiments and experiments obtained from taking limits in the sense of distributions. The latter are always weaker than the former. The situation where the two limits are equivalent deserves special mention. It is described here in terms of "stable" and "distinguished" sequences of statistics. It is shown that a sequence is distinguished only if one can approximately recover from it the likelihood ratios, and that in a nearly continuous manner. (See Theorem 1, and Proposition 2, Section 3). Theorem 2, Section 3 says that a sequence $\{T_n\}$ of statistics is "stable" if and only if any other sequence $\{T_n'\}$, such that $T_n - T_n'$ tends to zero in probability, is also asymptotically equivalent to T_n for all statistical purposes. The results are improvements on statements given in Le Cam [1972].

Section 4 deals with a form of what is sometimes called the Hájek–Le Cam asymptotic minimax theorem. The name comes from the works of Le Cam [1953], Hájek [1972], and Le Cam [1972]. Hájek's paper of 1972 contains other results on local asymptotic admissibility. A fairly general form of the results is given in Section 5, Theorem 1. This is adapted from Le Cam [1979].

Chapter 8 deals with some properties related to invariance of distributions or experiments under the operation of transformation groups. We have used a form of the Markov–Kakutani fixed point theorem obtainable from the arguments of F. Eberlein [1949]. For other results on almost invariant means or amenable groups, see Greenleaf [1969].

The Markov–Kakutani result yields easily the existence of what could be called "almost invariant Markov kernels." To make them actually invariant, we rely on a massive sledgehammer: locally compact groups admit liftings that commute with the group shifts (A. and C. Ionescu Tulcea [1967]). From this one easily obtains a form of the celebrated Hájek convolution theorem, from Hájek [1970], and some results of Torgersen [1972].

Section 4 of Chapter 8 deals with a peculiar phenomenon: In many situations statisticians pass to the limit for parameter spaces or statistics that are "renormalized" by multiplying by sequences of numbers (or matrices) that tend to infinity. Think, for instance, of the familiar $\sqrt{n}(\hat{\theta}_n - \theta)$. The peculiar phenomenon is that such renormalizations automatically imply certain invariance properties for the limits. We do this here for experiments and for limiting distributions. For a different account see Jeganathan [1982].

Section 5, Chapter 8 deals with a particular matter that arises from the above remark: If an experiment is invariant under shift and if it is an exponential family, what can it be? A partial characterization is given. It covers in particular the so-called mixed normal experiments, but includes certain other exponential families already described by E. Dynkin [1951].

Section 6 returns to invariance in general and gives a brief account of the Hunt–Stein theorem and of the Hall–Ghosh–Wijsman–Stein theorem [Hall-Ghosh-Wijsman [1965]), together with a variation by Torgersen [1976].

Chapter 9 is about infinitely divisible experiments. They occur readily as approximations, or limits, for experiments involving many independent observations. The two basic types are the Gaussian shift experiments and the experiments in which one observes Poisson processes. Every infinitely divisible experiment can be obtained from direct products of Gaussian and Poisson experiments. The Poisson experiments are characterized in Section 4, Lemmas 3 and 4. Their limits may involve Gaussian factors. They are described in Proposition 1. Section 5 gives a very simple central limit theorem showing that, for a finite parameter set, direct products of experiments that are individually not too informative can be approximated by Poisson experiments. This is extended in Proposition 4 to certain cases where the parameter space is infinite. Examples show that the result is not valid without restrictions.

Chapter 10 is about asymptotically Gaussian experiments. It gives a variety of necessary and sufficient conditions for convergence of a sequence of experiments to a Gaussian one. The natural parameterization for Gaussian experiments is a Hilbert space. Most of the results are written in that framework. The results were written many years ago. Many even preceded W. Moussatat's thesis [1976]. Since that time many authors, e.g., Millar [1979] and Beran [1980], have found the Hilbert space approach convenient and useful in nonparametric situations.

Theorem 1 is a basic asymptotic sufficiency result. Section 3 describes a fairly general framework that I met in many applications. A particular feature of it includes the possibility of replacing a parameter space by a space that is "tangent" to it. I used this since the mid-1950s without any references to the literature. That was so because, after a valiant effort, we did not find any.

However, W. Moussatat pointed out that the same idea occurs in a paper of H. Chernoff [1954]. It is hard to conceive that I had not heard Chernoff talk about it, but I certainly did not recollect it. Section 3 ends with some results on what has to be added to convergence of *distributions* to make experiments converge to Gaussian ones. It partially reproduces results of Le Cam [1977]. The gist of it is that if one has asymptotically normally distributed random vectors that behave as distinguished statistics, then the random part of the logarithms of likelihood ratios must be closely approximated by *linear* functions of these vectors.

Section 4 has to do with convergence in distribution to the standard Gaussian cylinder measure of a Hilbert space. Lemma 6 and Proposition 3 relate that to approximations of experiments indexed by the full Hilbert space by other experiments with a finite dimensional parameter space.

Section 5 is a digression intended mostly to show that for many stochastic processes one can obtain limit Gaussian experiments. It uses certain martingale limit theorems as suggested in Billingsley [1961] and Le Cam [1974]. For more recent results see P. Jeganathan [1981], Hall and C. Heyde [1980], and Greenwood and Shiryayev [1985].

Section 6 returns to asymptotically Gaussian experiments indexed by a Hilbert space. It gives a variety of asymptotic admissibility (or minimaxity) results derivable from a basic result of T.W. Anderson [1955]. Most authors prefer to Radonify the standard Gaussian cylinder measure of the Hilbert space by plunging it into a suitable Banach space completion. This is the same kind of thing as replacing white noise by a Wiener process. One can show that this does not change at all the standard Gaussian shift experiment of the Hilbert space. To obtain results about subconvex loss functions l with finite minimum risk, we have used instead an extension of the domain of definition of the Gaussian cylinder measure. An extension of this kind seems to have been used by Skohorod [1974]. One of the statements (Proposition 5 and its Corollary) is that sets of the type $B = \{w: l(w) \leq \alpha\}$ must have polars that are G.B. sets in the sense of Dudley [1967].

The arguments of Chapter 10 yield a number of results applicable in familiar cases to asymptotics in shrinking neighborhoods of a fixed parameter value. They are of little practical importance in cases where one desires to estimate the value of a parameter. For that, one needs results of a more global character. A particular set of results of that nature is given in Chapter 11. It relies on the fact that, given an already well behaved auxiliary estimate, one can work in a neighborhood of the estimated value and obtain more refined estimates that are asymptotically sufficient.

Any statistician who has at some time tried to get an approximation to a maximum likelihood estimate through Newton's method will understand immediately the spirit of that chapter. The technique used is the basis of the assertions in Le Cam [1956] and more generally in Le Cam [1960]. Unfortunately every time I wrote the necessary theory in what was supposed to be a definitive form, someone came along with a problem in which my conditions were not quite satisfied. Thus it appeared necessary to write down a rather general version of the theory. The version to be found here covers *every case* in which finite dimensional Gaussian shift approximations exist. It can be specialized to cases where appropriate Euclidean parameterizations are assumed. It also covers parameterizations through differentiable manifolds.

A previous version used tangent spaces, as in Chapter 10, Section 3. One of my friends complained that "Tu l'as écrit exprès pour qu'on ne puisse pas le lire." That was not true. However, that version was unable to cope with a very simple problem on logistic models brought to me by Clare Mahan [1979]. The new version is hopefully not subject to such difficulties.

The basic construction involves certain quadratic expressions that could be used as approximations for the logarithms of likelihood ratios of Gaussian experiments. In applications, the matrices used in these quadratics are often random, thus we have allowed that. There are cases in which the random character of the matrices is very essential; the construction still works.

Thus, the so-called "locally asymptotically mixed normal" families come under the scope of the method. Even though the chapter starts with that degree of generality, what it aims for is only the locally asymptotically Gaussian case.

Section 2 gives a few supplementary explanations on the why and how of the framework used here. Section 3 describes the items to be used in the construction and the construction itself. The basic construction assumes that certain quadratic forms are *available*. An alternate form uses the parallelogram identities to *construct* the necessary quadratic forms.

Section 4 gives a series of definitions relating to approximations of logarithms of likelihood ratios by quadratic expressions. The point is that the construction of Section 3 fits quadratics to these logarithms on a small set. One wants the approximations to hold on larger sets. Section 5 records such an approximability result in the form of Theorem 1 and a variant, Theorem 2, using quadratics obtained by the mid-point method. One of the conditions placed on the items used in the construction is a cardinality restriction on the range of the auxiliary estimates. It is shown that if one has estimates that converge at the right rate one can discretize them to satisfy the required restriction.

Section 6 specializes the results of Section 5 to the asymptotically Gaussian situation. Besides the basic approximability requirement (G1) of Definition 1, it uses various conditions that bear on the items used in the construction, as in Section 5. It is shown later on (Lemma 4 and Proposition 5) that, under (G1) and a few other restrictions, the basic sets and quadratics of the construction can themselves be constructed. Thus the only very essential assumption to be made otherwise is the existence of good auxiliary estimates. Even though the asymptotically finite dimensional case is the only one for which entirely satisfactory results exist, we have indicated some possibilities for precompact infinite dimensional situations. In the asymptotically Gaussian case one needs to work with random quadratic forms and with nonrandom ones. There are many possibilities depending on what needs to be achieved. We have indicated the possibility of constructing several families of quadratic forms, with different properties. Theorem 1 gives a basic approximation property. Proposition 1 records some of the usual implications for convergence of distributions. Theorem 2 is an aside that may be useful to deal with nuisance parameters. Theorem 3 is a basic result on asymptotic sufficiency, imitated from Le Cam [1956] and Le Cam [1960]. These theorems do not depend very much on the local structure of the underlying parameter spaces. The assumptions most often used in the literature imply that, locally, the structure of the parameter space is very much like that of an entire Euclidean space. This leads to Definition 8 and some of its consequences.

Section 7 specializes the results of Section 6 to those cases commonly referred to as LAN in the literature. They assume a Euclidean structure for the parameter space and special properties of the quadratic forms and norms used.

Section 8 is aimed at showing that one can effectively reduce the approximately Gaussian case to the heteroschedastic Gaussian one by small distortions. Results of this nature occur in Wald [1943] and Le Cam [1956]. Michel and Pfanzagl [1970] show that one can approximate the experiments by others where the likelihood ratios have exactly the Gaussian form, except that they are multiplied by functions of θ that tend to unity. Section 9 records optimality properties for the standard tests used in the Gaussian case. Then it shows how to translate them for the asymptotic framework. It is imitated from Wald [1943]. It also discusses very briefly the locally asymptotically mixed normal situation.

Section 10 is a digression on chi-square, minimum chi-square estimates, and similar topics. It is taken mostly from my class notes of 1959 and before. However, we have added some techniques of Rao and Robson [1974]. Some additional information can be found in Dudley [1976] and in Le Cam, Mahan, and Singh [1983].

Chapter 12 is about convergence properties for posterior distributions. Throughout, the attitude is to look at joint distributions and not at what happens at particular values of the parameter θ. Section 2 deals with some simple inequalities relating integrated L_1-distances between posterior distributions to distances between joint distributions. It also contains some results about the possibility of "localization" of the arguments. It does not cover the results of G. Steck [1957] that rely on equicontinuity assumptions. Section 3 is a rewrite of the first part of Le Cam [1958]. It tries to say that Bayes estimates will converge at the maximum possible rate. Section 4 deals with situations where the posterior distributions are asymptotically Gaussian. A result of the same general occurs in Laplace [1820].

Several of the results collected here were taken from an unpublished paper of Le Cam [1968]. Proposition 5 occurs in Le Cam [1974] but the proof given there is not correct. For further results see Jeganathan [1980], and Basawa and B.L.S. Prakasa Rao [1980].