Statistical Science
2019, Vol. 34, No. 4, 621–634
https://doi.org/10.1214/19-STS717

Larry Brown's Contributions to Parametric Inference, Decision Theory and Foundations: A Survey

James O. Berger and Anirban DasGupta

Abstract. This article gives a panoramic survey of the general area of parametric statistical inference, decision theory and foundations of statistics for the period 1965–2010 through the lens of Larry Brown's contributions to varied aspects of this massive area. The article goes over sufficiency, shrinkage estimation, admissibility, minimaxity, complete class theorems, estimated confidence, conditional confidence procedures, Edgeworth and higher order asymptotic expansions, variational Bayes, Stein's SURE, differential inequalities, geometrization of convergence rates, asymptotic equivalence, aspects of empirical process theory, inference after model selection, unified frequentist and Bayesian testing, and Wald's sequential theory. A reasonably comprehensive bibliography is provided.

Key words and phrases: Admissibility, ancillary, asymptotic equivalence, Bayes, conditional confidence, differential inequality, Edgeworth expansions, estimated confidence, minimax, sequential, shrinkage, sufficiency.

1. INTRODUCTION

In his expository article "An essay on statistical decision theory" (2000), Brown gives a panoramic picture of the historic evolution of decision theory from the times of Neyman and Pearson to the beginning of the century, and argues with numerous examples, applications, and demonstrations that *the spirit of decision theory is pervasive in contemporary statistical research* (italics the present authors'). For a span of over half a century, Larry Brown made cardinal and characteristically original contributions to most of these changing aspects of decision theory.

Finite sample optimality theory was the dominating theme of theoretical statistics from about the mid forties to the mid sixties. Asymptotics have such a central place in the development of statistical methods because it is rare that exact finite sample calculations can be done in closed form. But when they can be done, the optimality results are generally very beautiful. This paramount beauty of the optimality theory of parametric statistics has repeatedly come out in some of its most delicate and elegant form in Brown's work. It has been deep, original, peerless and difficult. We will summarize Brown's most influential contributions to this dynamic and fascinating evolution of decision theory and the far-flung decision theoretic spirit.

Brown also continually strove to understand foundational issues in statistics. While not a Bayesian, he was continually trying to understand matters from the Bayesian side, as a way to enhance frequentist statistics. Perhaps the culmination of this was his foundational work on the conditional frequentist perspective, which spanned his entire career and demonstrated his endless attempt to understand statistical foundations. Since the article is arranged in a chronological, rather than subject matter order, these foundational advances are listed at the end of each section.

The article of Tony Cai in this volume details Brown's work on function estimation and other infinite dimensional problems. We do not cover them in this article.

James O. Berger is Arts and Sciences Professor, Department of Statistical Science, Duke University, Durham, North Carolina, USA (e-mail: berger@stat.duke.edu). Anirban DasGupta is Professor, Department of Statistics, Purdue University, West Lafayette, Indiana, USA (e-mail: dasgupta@purdue.edu).

2. THE SIXTIES: SHRINKAGE, COVERAGE PROBABILITIES, SUFFICIENCY

Brown was advised by algebraist Robert Dilworth to work with Jack Kiefer at Cornell for his doctoral work. He took Dilworth's advice. Stein's unexpected result on the inadmissibility of the sample mean, which is the MLE as well as the best invariant estimator, for estimating three or more normal means was well known when Brown started working with Kiefer. For more general location parameter distributions, the sample mean is neither the MLE nor the best invariant estimator (Kagan, Linnik and Rao, 1973). By that time, Stein had started to work (Stein, 1959, James and Stein, 1961) on generalizing both the normal case results as well as the results of Blackwell (Blackwell, 1951) on admissibility of the best invariant estimator of a location parameter in the finite discrete case. Brown's association with Jack Kiefer as his doctoral advisor led to the 1966 work (Brown, 1966) that set an example for demonstration of universality and the absoluteness of a phenomenon, that almost like a law of nature, for essentially any location parameter situation and with any loss function, pathologies aside, there is a dichotomy between admissibility and inadmissibility of the best invariant estimator in dimensions two or less and dimensions three or more. The conditions in Brown (1966) are moment conditions, as in Stein (1959), where it was shown that the existence of a third absolute moment is sufficient for minimaxity of the best invariant estimator. Interestingly, similar moment conditions are also necessary to prove the admissibility of the best invariant test for a location parameter; the principal reference is Lehmann and Stein (1953). The article of Iain Johnstone in this volume details the technical conditions and the main theorems of Brown (1966), the pathological counterexamples, as well as his principal motivation in presenting his results in such sweeping generality. The results in Brown (1966) can be applied to scale parameter problems via the usual logarithmic transformation; for example, it can be proved that in the multivariate normal setting, the best scale invariant estimator of $|\Sigma|$ is admissible under scaled quadratic loss (DasGupta, 1983). As for estimation of the mean vector in the multivariate normal setting when the covariance matrix Σ is unknown, but there is available an independent Wishart matrix to estimate it, estimators that dominate the best invariant estimator of the mean were given in Berger et al. (1977).

While the Brown (1966) paper was clearly influenced by Brown's association with Jack Kiefer, we can only speculate that another extremely difficult and fundamentally important paper on sufficiency, Brown (1964), was influenced by Eugene Dynkin. Through its well-known connections to the Rao–Blackwell and Lehmann–Scheffe theorems, sufficiency is an integral component of the study of optimality for finite n, and is, as such, a part of decision theory. If $\{P, P \in \mathcal{P}\}$ is a family of probability measures on some measurable space $(\Omega, \mathcal{A})$, a sub σ-field $\mathcal{B}$ of $\mathcal{A}$ is sufficient if for $\forall P \in \mathcal{P}$, and for each set $A \in \mathcal{A}$, there exists a single $\mathcal{B}$-measurable function g_A such that $g_A = E(I_A|\mathcal{B})$, a.e. P. In common usage, we generally state it as saying that if T is a mapping from the original space into some other measurable space $(\Omega', \mathcal{A}')$, then it is a sufficient statistic if $\mathcal{B} = \mathcal{B}_T = T^{-1}(\mathcal{A}')$ is a sufficient sub σ-field of $\mathcal{A}$. Fisher (1923) stated and variously, Halmos and Savage (1949), Bahadur (1954), and others, gave the factorization theorem, which is the standard result we present in a classroom situation. If the family $\mathcal{P}$ is parametrizable by a one dimensional Euclidean parameter θ, it is rare that a single real statistic $\phi(X_1, X_2, \dots, X_n)$ is sufficient for every n. It was stated in Lehmann (1959) and was very much a part of the statistical folklore that this will force the family $\mathcal{P}$ to be a one parameter exponential family under regularity conditions. In Brown (1964) and essentially simultaneously in Barankin and Maitra (1963), somewhat different versions of this fundamental result are proved. Brown (1964) gives two results establishing this implication, one global and another local, depending on what one is willing to assume about the shape of the sufficient statistic as a function. The global theorem says the following. Suppose $\{f(x|\theta)\}$ is a one parameter family of densities on an interval I (possibly unbounded) in the real line, with each $f(x|\theta)$ being mutually absolutely continuous with respect to λ, Lebesgue measure on I. Suppose a statistic $\phi(X_1, X_2, \dots, X_n)$ is such that for every given $x_3, \dots, x_n$, $\phi(x_1, x_2, x_3, \dots, x_n)$ is jointly continuous in (x_1, x_2) for $x_1 \in I_0$ and $x_2 \in I$ for some set I_0 of a positive Lebesgue measure. If $\phi(X_1, X_2, \dots, X_n)$ is sufficient for all n, then the family $\{f(x|\theta)\}$ must be a one parameter exponential family.

In the interesting but special case when the statistic is a sample mean $\sum_{i=1}^n \psi(X_i)$, where the function ψ may not be continuous everywhere, Brown (1964) shows that around each continuity point x_0 of ψ, the family $\{f(x|\theta)\}$ is locally a one parameter exponential family, provided ψ is nondecreasing.

Various counterexamples, some new, and others explaining and interpreting counterexamples of Dynkin (Dynkin, 1961) are given in the Brown (1964) paper.

Brown (2000) gives a historical account of how Neyman viewed confidence interval procedures and confidence levels as a part of decision theory, and Brown agrees with this position. But Brown differed from Neyman in one important foundational aspect; throughout his career he was concerned with issues of conditioning within frequentist statistics, whereas Neyman rarely mentioned the issue.

Brown's first work on conditional frequentist inference was Brown (1967), which considered the widely used t-test. In particular, he considered $N(\mu, \sigma^2)$ data, $x_1, \ldots, x_n$, with μ and σ^2 unknown, with the desire to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ via the t-test at Type I error level α, having corresponding rejection region K_{α,μ_0}. Brown considered conditioning on the set of possible data $C = \{(x_1, \ldots, x_n) : t \leq c\}$, where t is the t-statistic, and showed that $\alpha(C)$, the Type I error conditional on C, satisfies

$$\alpha(C) = \Pr(K_{\alpha,\mu_0} \mid C, \mu_0, \sigma^2) < a < \alpha,$$

for some constant a and any values of μ_0 and σ^2. (The proof was a careful and difficult geometric calculation.)

As an example, Brown shows that if $n = 2$, $\alpha = 0.5$, and $c = 1.414$, then $\alpha(C) \leq 1/3$. It is perplexing that one can find sets which, when conditioned upon, yield Type I errors smaller than 1/3 for all values of the parameters, when the unconditional coverage is 1/2. But Brown admits that this is a puzzle, saying "In view of the well-known optimum properties of the t-test it is not clear that the results of this note can possibly lead to any practically useful new procedures. (It is not even clear that any remotely reasonable test procedures exist for this problem which do not have conditional properties similar to those described here.)" We do not believe he is referring to the technical issues inherent in this problem, or to issues of generalization, but rather to the difficulty of rationally determining appropriate conditioning sets (in advance) for this problem. Thus we see Brown beginning to struggle with the issue of frequentist conditioning, in a technically very challenging problem, but a situation for which philosophical clarity is lacking.

3. THE SEVENTIES: ADMISSIBILITY IN ESTIMATION AND SEQUENTIAL TESTING, FOUNDATIONS

The famous paper connecting admissibility, recurrence of diffusions, and boundary value problems was written in 1971. Although the paper was published when he was at Cornell, it was in the works for some time before then, and he had the basic structure already in place while he was still at Berkeley. The paper of Iain Johnstone in this volume discusses this paper in detail. The '71 paper is famous for at least three reasons. Potential theory is connected to recurrence of diffusions. Brown (1971) imports a shocking new connection, statistical admissibility, by all means apparently completely unrelated, and makes a triangle with three vertices. Second, it gives fully verifiable sufficient conditions for admissibility, which have been successfully employed in numerous Bayesian contexts, for example, Berger (1976), Berger and Strawderman (1996). Third, the paper is an analytical masterpiece. The connections demonstrated in the 1971 paper in the normal case were examined and established in various other settings, for example, Poisson means (Johnstone, 1984, Johnstone and Lalley, 1984). Eaton (1992) gave a very interesting equivalence in all *quadratically regular* problems between almost admissibility of formal Bayes procedures with respect to σ-finite improper prior measures ν and the recurrence of a symmetric Markov chain. It is interesting that Eaton's Markov chain is on the *parameter space*. If ν supports every set with a nonempty interior and if all procedures with a finite risk function have a continuous risk function, then the almost part of the admissibility connection drops. Exponential families are good examples. The connections between Brown (1971) and Eaton (1992) are not yet well understood. Another difficult and fundamental paper in this same period is Brown and Purves (1973), which gives necessary or sufficient conditions for the existence of a Borel measurable function $\phi(x)$ such that for all x, $f(x, \phi(x)) = \inf_y f(x, y)$, where f is a Borel measurable function on a product space $X \times Y$ in $\mathcal{R}^2$. The conditions are used to give an example where pointwise minimization of the posterior expected loss does not give a measurable Bayes estimator.

While at Rutgers University, Brown wrote a sequence of papers characterizing sequential admissible tests and more or less characterizing, in particular, SPRTs that are admissible in terms of their stopping boundaries. In the i.i.d. (or independent) situation, Wald's SPRT (Wald, 1945) stops and rejects if the likelihood ratio is too large, and stops and accepts if the likelihood ratio is too small, and continues sampling otherwise. This works out to a test of the form for which the continuation region is $a_n \leq S_n \leq b_n$, where $S_n, n \geq 1$ is a sequence of partial sums, and a_n, b_n are suitable real sequences. Suppose the risk of a test is measured by the probability of taking the wrong action plus a constant c times the expected sample size of the test before termination; c is thought of as the cost of sampling one unit. Sobel (1953) and Brown, Cohen and Strawderman (1979) had established that the so called *monotone* tests form an essentially complete class in this risk structure. Essentially, a monotone test accepts with probability zero at any stage that the likelihood ratio is too large, rejects with probability zero if the likelihood ratio is too small, and continues sampling with probability zero if the likelihood ratio is too small or too large. It is shown in Brown, Cohen and Samuel-Cahn (1983) that the Bayes tests, a much smaller subclass of the class of monotone tests, forms a complete class. It is important to note the difference of this phenomenon from admissibility in the fixed sample size setting.

It stands to reason that if cost of sampling is too high, one should not sample for too long. Indeed, in the exponential family setting, Brown and Cohen (1981) show that any test with $\limsup(b_n - a_n) = \infty$ is inadmissible. Thus, admissible tests must have a bounded continuation region. Brown, Cohen and Samuel-Cahn (1983) show that all admissible tests must satisfy $b_n - a_n \le \bar{b}(c)$ for a suitable function $\bar{b}(c) = O(c^{-1/2})$ as $c \to 0$. In particular, all admissible SPRTs are characterized.

During this time Brown again returned to the issue of conditioning in frequentist statistics, in the paper Brown (1978). This was motivated by Kiefer (1977), which sought to develop a formal theory of frequentist conditioning through decision theory. A key component of Kiefer's work was that it allowed statements of conditional confidence level of the form $C(S, \theta)$, which depended on the unknown parameter θ as well as the conditioning set of possible data S. Brown viewed such statements as being of limited practical usefulness, and suggested using, instead, $C^*(S) = \inf_\theta C(S, \theta)$. He considered admissibility and minimaxity (of three types) for decision problems involving such conditional confidence statements. For certain situations where θ is dichotomous, he solved these decision problems, obtaining the optimal conditional confidence procedures. Brown did not follow this work up, however, perhaps realizing that the case of dichotomous θ was so difficult that more complex situations would not be tractable.

4. FROM PERSUASIVE TO CONFIRMED: THE HEURISTICS AND DIFFERENTIAL INEQUALITIES PAPERS

The two papers Brown (1979) and Brown (1988) are thematically connected. The latter is directly connected to Stein's unbiased risk estimate (SURE, Stein (1981)), and the first one is indirectly connected to the SURE theme. The paper of Iain Johnstone in this volume gives the technical details related to these two papers. The common premise of both of these papers is to approximate the difference in the exact risk functions of two procedures $X + g(X)$ and $X + g(X) + \lambda(X)$ by a differential operator $\mathcal{R}_0(\lambda)$ or the expectation of it. Sometimes the operator is defined on the parameter space, and sometimes on the sample space. $X + g(X)$ will be inadmissible if the operator can be made uniformly nonpositive by a suitable choice of a function in some function class. Treating $\mathcal{R}_0$ itself as the risk, a variational formula for Bayes risk is derived, and the form of the Bayes procedure is obtained as the minimizer of this functional. The existence of a solution to $\mathcal{R}_0(\lambda) \le 0$ is then connected to whether or not the function g is one of the above described Bayes forms. In other words, a starting estimator is conjectured to be admissible if it is admissible with respect to the $\mathcal{R}_0$ risk only if it is of the Bayes form for the $\mathcal{R}_0$ risk. In some special but important cases, the implication is if and only if (see, e.g., (7.2) in Brown, 1988). See the paper of Iain Johnstone in this volume.

The approaches laid out in both papers have the purpose of taking a hard new problem, and use Brown (1979) or Brown (1988) to predict the admissibility status of a given procedure or a given class of procedures. One then turns the prediction into a proof by some suitable direct argument. If this seems too optimistic, it is not. Brown gives a collection of examples of a wide variety where this two step agenda, predict and confirm, succeeds. Some of Brown's examples are stated in Iain Johnstone's paper in this volume.

5. THE EIGHTIES: FOUNDATIONS, BAYESIAN AVENUES, COMPLETE CLASS THEOREMS

The Stein–Blyth method (see Diaconis and Stein, 1982, Berger, 1985, Rukhin, 1995) having been the principal tool for proving the admissibility of a specific estimator, anyone interested in admissibility automatically uses priors as an essential tool. Furthermore, there is often a link between all admissible estimators and all Bayes or extended Bayes estimators. In Brown (1966), dominating estimators, are not arrived at from an intrinsically Bayesian spirit. They derive more from the spirit of pure shrinkage, without relating shrinkage to Bayes (or empirical Bayes, etc.). In Brown (1971), greater technical use of priors is made. Furthermore, Brown (1971) derives results directly pertinent to Bayes, for example, the Brown identity for the Bayes risk. In contrast, in the eighties, we see the evidence of a growing natural interest in Bayes as a practical tool in a number of papers of Brown. One example is the discussion paper Brown (1982) on robust Bayesian analysis. In a number of other very influential papers, Bayes is still used as a tool. Primary among them is the highly original Brown and Hwang (1982) paper, and the Brown (1981) paper.

Brown (1981) gives a method to completely characterize all admissible procedures in any problem with a finite sample space, with suitable smoothness conditions on the loss functions etc. The technique is to write a sequence of restrictions of the original sample space, together with an associated modification for the corresponding parameter space. If a procedure is Bayes with respect to some prior in each subproblem, it is called totally Bayes. The totally Bayes procedures form a complete class in the original problem, and any unique (with respect to the sequence of priors) totally Bayes procedure is admissible in the original problem. This paper extends the stepwise Bayes algorithm in Hsuan (1979) and is concretely usable to prove that a certain procedure in a finite sample space problem is admissible. One example is Meeden et al. (1989). In addition, it proves the admissibility of the MLE in the multinomial problem, which was previously proved by using the Cramér–Rao inequality (Olkin and Sobel, 1979). It also reproduces the results in Johnson (1971) and Skibinsky and Rukhin (1989). The paper of Iain Johnstone in this volume lists and explains the other complete class theorems proved by Brown and various coauthors in different papers written in this same period.

Brown and Hwang (1982) is a remarkably insightful paper for several reasons. First, it, for the first time, brings out the role of the asymptotic flatness of the prior density in determining the admissibility status of a generalized Bayes estimator. Second, it gives, for the first time, a Blyth-based proof of Karlin's theorem (Karlin, 1958) on admissibility of linear estimators of the mean in a one parameter exponential family. Third, it gives a unified proof of admissibility of subclasses of dominating estimators in the multiparameter case, for example, the admissible ones among the Clevenson–Zidek estimators (Clevenson and Zidek, 1975) of multiple Poisson means. Fourth, the paper succeeds in applying Blyth's theorem by using a sequence of compactly supported priors.

In the multiparameter exponential family setting with $\theta_{p\times 1}$ as the natural parameter, suppose $\delta_g(X)$ is the generalized Bayes estimator of the mean parameter for a prior density $g(\theta)$ with squared error as loss. Take g to be a continuous function so that it has a finite maximum on all compact sets. Suppose g satisfies

$$\int_{\|\theta\|>2} \frac{g(\theta)}{\|\theta\|^2 \log^2(\|\theta\|)}\, d\theta < \infty,$$

and that g has finite Fisher information. The leading theorem in Brown and Hwang (1982) says that then $\delta_g(X)$ is admissible. It is important to note that in the above, $\|\theta\|$ denotes the Euclidean norm, for in certain papers that followed Brown and Hwang (1982), the definition of the norm needed to be changed. The approximating sequence required for applying Blyth's method is $g_n(\theta) = g(\theta)h_n^2(\theta)$, where

$$h_n(\theta) = 1 I_{\|\theta\|\le 1} + \left(1 - \frac{\log(\|\theta\|)}{\log n}\right) I_{1\le\|\theta\|\le n} + 0 I_{\|\theta\|\ge n}.$$

Thus, for each n, g_n is compactly supported. The method of Brown and Hwang (1982) is adapted *with a new norm* $\|\theta\|$ in DasGupta and Sinha (1984) to prove parallel admissibility results for generalized Bayes estimators of k linear combinations of the components of the mean vector for some k, typically smaller than p. The case $k = 1$ was previously studied in Cohen (1965) for the normal case using different methods.

The Brown identities (Brown, 1971) were put to significant use in Brown and Hwang (1982). It is interesting that the Brown identities also help answer certain natural questions in frequentist Bayesian asymptotics. In regular finite dimensional problems with smooth priors, under natural conditions on the support of the prior, the posterior mean will be asymptotically pointwise close to the MLE, and the two have the same asymptotic distribution. The Brown identities can be used to derive the asymptotic distribution of their difference. Thus, suppose $X_1, X_2, \ldots$ are i.i.d. $N(\theta, 1)$, and that θ has a prior density $g(\theta)$ which is three times differentiable, and that g'', $g^{(3)}$ are bounded. Let $\nu(\theta) = \frac{g'(\theta)}{g(\theta)}$. Let θ_0 be fixed (consider this as the *true value*), and assume that $\nu'(\theta_0) \neq 0$ (notice that $\nu(\theta_0)$ can be zero, for example, if $\theta_0 = 0$ and $g(\theta)$ is a symmetric unimodal function). Then, under P_{θ_0},

$$\sqrt{n}\big[n\big(E(\theta|X_1,\ldots,X_n) - \bar{X}\big) - \nu(\theta_0)\big] \xrightarrow{\mathcal{L}} N\big(0, [\nu'(\theta_0)]^2\big)$$

(see, e.g., DasGupta, 2008).

At the end of the eighties, Brown again returned to the foundational issue of frequentist conditioning, work that appeared in Brown (1990) as part of his earlier Wald Lectures.

A widely accepted "principle" of statistics is that, if there is an *ancillary* statistic $S(x)$ (i.e., a statistic whose distribution does not depend on the unknown parameter θ), then one should perform statistical inference conditional on $S(x)$. There did, however, exist famous examples where conditioning on ancillary statistics results in inadmissible procedures, if one uses apparently natural frequentist criteria.

EXAMPLE. In the famous Cox example, depending on the flip of a fair coin, one observes either $X \sim N(\theta, 1)$ or $Y \sim N(\theta, 4)$, and wishes to test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. If one were to condition on the ancillary coin flip and use, say, an $\alpha = 0.05$ level test for either X or Y, one has a testing procedure with lower power than optimal. But, conditionally, one *can recognize there is a problem*; for example, the conditional $\alpha = 0.05$ tests are Bayes with respect to different prior distributions on θ which implies their unconditional inadmissibility.

The brilliance of Brown's paper is that he found a host of fascinating examples where conditioning on ancillary statistics led to unconditional inadmissibility, and one could not conditionally detect the problem (e.g., the same priors are utilized always), the simplest of which is the following.

EXAMPLE. For $i = 1, \ldots, n$, we are interested in estimating α, based on independent data $Y_i = \alpha + X_i\beta + \epsilon_i, \epsilon_i \sim N(0, 1)$, where the covariates X_i arise from a normal distribution with mean zero and given covariance matrix. The covariates X_i are ancillary, so one "should" condition on them in performing the inference. But Brown shows that:

- $\hat{\alpha}$, the least squares estimate of α (which does effectively condition on the covariates), is unconditionally inadmissible under squared error loss, if the dimension of β is 2 or more;
- $\hat{\alpha}$ is admissible if the X_i are fixed;
- $\hat{\alpha}$ would arise as the Bayes estimate using the improper prior density $\pi(\alpha, \beta) = 1$, no matter what the X_i are, so there is no conditional hint of an inadmissibility problem.

There are a lot of mysteries here, and no one (including a stellar group of discussants of the paper) has come up with an explanation or solution that has been generally accepted as being satisfactory to both conditionalists and frequentists.

6. THE EXPONENTIAL FAMILY MONOGRAPH

The exponential family monograph (Brown, 1986) is a major contribution toward statistical, geometric and probabilistic unification of scattered results on common statistical distributions under the common umbrella of the regular and curved exponential families. The monograph can be used to give a semester's course or longer on the standard optimality theory of parametric statistical inference, covering maximum likelihood estimation, UMVUEs, Bayes, minimax and admissible procedures, UMP tests, and similar, unbiased, and invariant tests. A particular asset of the monograph is the exercises, many of which are nonstandard, but interesting, and apply the theory to popular models or problems.

The first three chapters describe in detail and with proofs essentially all the standard analytic and probabilistic properties of exponential families. The curved exponential families are given a unified treatment through consideration of a property called *steepness*. A number of examples are nonstandard, and so are a number of the proofs. The approach taken is to unify the analytic and probabilistic properties with the geometry of the family and the parametrization. Apart from parametrization by either the natural parameters or the mean parameters, a mixed parametrization is introduced and analyzed. The standard theory of testing of composite hypotheses is also presented here. These are the problems where UMP tests do not exist, and reductions to unbiasedness, Neyman structure, invariance, or almost invariance are needed. It is good to see that the Hunt–Stein theorem, very hard to find in texts, is covered.

Chapter 4 forms the core of standard decision theory in the exponential families, introducing SURE, variational identities known as the *Brown identities*, methods for proving admissibility, and generalized Bayes estimators and their analytic properties. Chapter 5 and Chapter 6 cover the likelihood theory in a nonstandard way. Existence and uniqueness of the MLE is treated as a problem in convex duality and augmentation of the original family to aggregate families in which the set of MLEs is always nonempty. Relationships to the elements of the set of MLEs to the minimal sufficient statistic is examined theoretically and by examples.

Chapter 7 treats the exponential decay of probabilities of a sequence of sets as the distance of the set from the mean increases suitably, or when the set is fixed, but the mean parameter is moving away from it suitably. These are analyzed as large deviation problems.

The exponential decay is then connected to powers of tests in various problems.

The appendix gives a complete treatment of limits of Bayes procedures and the necessity of admissible procedures being such limits.

7. THE NINETIES: MINIMAX AND BAYES RISKS LOWER BOUNDS, FOUNDATIONS, NOVEL CRAMÉR–RAO USES, ASYMPTOTIC EQUIVALENCE

Brown used the Cramér–Rao inequality in ingenious ways in a number of different problems to arrive at useful results at the end. Brown and Farrell (1990) made clever uses of the Cramér–Rao inequality to obtain calculable lower bounds on the asymptotic minimax risk for estimating the value at a given point of a locally Lipshitz density on the real line. One major family of problems for which Brown made novel use of the Cramér–Rao inequality is to obtain lower bounds on Bayes risks for general (possibly smooth) priors, and hence lower bounds on the minimax risk by applying the Bayes risk bounds to least favorable (or nearly so) priors. Among the most well known is Brown and Gajek (1990). We will describe one other unrelated problem where Brown made a very novel use of the Cramér–Rao inequality following our description of the Brown and Gajek (1990) work.

Let $X \sim f(x|\theta), \theta \in I$, a possibly unbounded interval in the real line, $\delta(X)$ any estimator with a finite variance and bias $b(\theta)$, $g(\theta)$ a prior density and an absolutely continuous function. Suppose the family has finite nonzero Fisher information $I(\theta) = V^{-1}(\theta)$. Let

$$C = \int (Vg)(\theta)\, d\theta, \qquad D = \int \frac{[(Vg)'(\theta)]^2}{g(\theta)}\, d\theta.$$

Then, for estimation with squared error loss, the Bayes risk of δ with respect to g satisfies

$$B(g, \delta) \geq \frac{C^2}{C + D} + Q,$$

where Q is a nonnegative real valued functional of $b(\theta)$ and its derivative $b'(\theta)$. In particular, $B(g, \delta) \geq \frac{C^2}{C+D}$, which is the Borovkov–Sakhanienko (1980) lower bound, proved using different methods. The Brown–Gajek bounds generalize to weighted squared error losses.

The lower bound is attained in exponential families with conjugate priors. For an interesting application, consider the Bickel–Levit prior for a bounded normal mean

$$g(\theta) = \frac{1}{m} \cos^2\left(\frac{\pi\theta}{2m}\right), \quad |\theta| \leq m.$$

Then the Brown–Gajek bound produces

$$B(g) \geq \frac{m^2}{m^2 + \pi^2},$$

where $B(g)$ is the Bayes risk of the associated Bayes estimator; see Bickel (1981) for a treatment and asymptotic expansion of $B(g)$. In fact, the Brown–Gajek theorem gives a better closed form bound on $B(g)$. Other examples, including a binomial log-odds example are worked out in Brown and Gajek (1990).

Bayes risk lower bounds are also obtained in Brown et al. (2006) by using a new normal distribution expectation identity, which they call *the heat equation identity*. The paper of Iain Johnstone in this volume gives a discussion of some of the results in this paper. They derive a Bayes risk differential identity which can be used to derive Bayes risk lower bounds. Suppose $X \sim N(\mu, t)$ and μ has a prior distribution G. Let $m(x) = m_{t,G}(x)$ denote the marginal density of X and $r(t, G)$ the Bayes risk of the Bayes estimator under G. Then the article derives the identity $\frac{d}{dt} r(t, G) = 1 - 2tI(m) - t^2 \frac{d}{dt} I(m)$, where $I(m)$ denotes the Fisher information of the marginal. On integrating this differential identity, one obtains Bayes risk lower bounds.

Minimax lower bounds using the Cramér–Rao inequality and other techniques (notably the hardest linear subproblem technique of Donoho, Liu and MacGibbon, 1990) are derived in several subsequent papers of Brown, including Brown and Low (1991).

The article Donoho, Liu and MacGibbon (1990) opened up a very new line of minimax decision theory problems, techniques, and connections to hitherto mostly unfamiliar applied mathematics constructs. Perhaps this article marked a noticeable shift in Brown's research efforts from admissibility to minimaxity, and indirectly, to more Bayesian ideas. Consider the problem of estimating with squared error loss an infinite dimensional normal mean with the model

$$X_i = \theta_i + \epsilon_i, \quad i = 1, 2, \ldots,$$

where ϵ_i are i.i.d. $N(0, \sigma^2)$ and the vector θ lies in the infinite dimensional hyperrectangle $\prod_{i=1}^{\infty} [-\tau_i, \tau_i]$. The exact minimax estimate or the minimax risk $R_N(\sigma)$ cannot be found in analytical form. It is a nonlinear estimate. In contrast, the linear minimax estimate is easily found to be

$$\hat{\theta}_{L,i} = \frac{1}{1 + \frac{\sigma^2}{\tau_i^2}} X_i.$$

The linear minimax risk $R_L(\sigma)$ is easily found by the coordinatewise sum of the corrdinate linear minimax risks. It is shown in Donoho, Liu and MacGibbon (1990) that $\frac{R_L(\sigma)}{R_N(\sigma)} \leq 1.25$. It should be mentioned that this result is partially computer aided. Donoho, Liu and MacGibbon (1990) then go on to show that the 1.25 upper bound on the inefficiency of the linear minimax estimate also holds for all quadratically convex parameter spaces, not just hyperrectangles. Thus, if the normal mean vector lies in an l_p ball with $p > 2$ or in an ellipsoid, the 1.25 upper bound still holds.

Brown and Liu (1993) develop lower bounds on the minimax risk in a signal plus noise model

$$x(t) = s_\theta(t) + \sigma W(t), \quad 0 \leq t \leq T,$$

where $0 \leq \theta \leq L$, the signal $s_\theta(t)$ is a step function $s_\theta(t) = SI_{0 \leq t-\theta \leq K}$, and $W(t)$ denotes a Wiener process on $[0, T]$. Interest is in minimax estimation of θ on the basis of the signal received $\{x(t)\}_{t=0}^{T}$. Lower bounds on the Bayes risk (which also give lower bounds on the minimax risk) are provided for a fully uniform prior as well as discrete priors supported at multiples of K. These lower bounds are derived by using a variety of techniques, including an often adopted two-point trick in minimax theory (see Tsybakov, 2004 or Korostelev and Korosteleva, 2011). In the interesting case where K is held fixed and $L \to \infty$, it is proved that the minimax risk is $\sim L^2/4$, and so is the Bayes risk of the discrete prior Bayes estimate. The convergence rate is pinned down by an asymptotic expansion. It is shown that the MLE is not minimax efficient and its efficiency is $3/4$.

Earlier, Brown (1982) made yet another novel use of the Cramér–Rao inequality to prove the i.i.d. case CLT for a subclass of random variables with finite variance. Thus, suppose $X_1, X_2, \ldots$ is an i.i.d. sequence with mean zero, unit variance, finite and nonzero Fisher information $I(X) = \int \frac{f'^2}{f}\,dx$, where $f(x)$, the density of X_1 is assumed absolutely continuous. A well-known variational property is that Fisher information is minimized at normal distributions; thus, $I(X) \geq 1$, with equality for a normal with variance one. It is shown in Brown (1982) that the sequence of Fisher informations of the normalized partial sums, $I(S_n)$, is a strictly decreasing sequence converging to one. This is used to prove that the integrals of enough functions also converge, and hence S_n converges weakly to $N(0, 1)$. Proofs of the CLT based on entropy, instead of Fisher information, were given in Barron (1986) and Johnson (2004).

Another elegant article in this period is Liu and Brown (1993). The article explains in a unified way why in many specific examples of nonparametric estimation or parametric estimation with a singularity, no unbiased estimators or unbiased estimators with a finite variance can exist. In particular, it is shown that if the second moment of an estimator sequence is Hellinger-continuous, then it cannot be locally unbiased at any singular point of the parameter space.

In the nineties, we also see the beginning of a large body of work on asymptotic equivalence of two sequences of decision problems in two different sequences of spaces, but with the same parameter space. Equivalence holds in a very wide sense. Risks or rates attainable in one problem are attainable in the equivalent problem. If a minimax procedure is obtainable in one problem, it leads to a minimax procedure in the equivalent problem.

Asymptotic equivalence is defined in terms of Le Cam's metric for the distance between two (sequences of) experiments (Le Cam, 1986), say, $\Delta(\mathcal{P}_{n,1}, \mathcal{P}_{n,2})$. Two sequences of experiments are equivalent if $\Delta(\mathcal{P}_{n,1}, \mathcal{P}_{n,2}) \to 0$ as $n \to \infty$. If two sequences of experiments are equivalent, then any risk profile attainable in one problem is also attainable in the other problem. Thus, given a sequence of decision rules $\delta_{n,1}$ in the first problem, there is a sequence of decision rules $\delta_{n,2}$ in the second problem such that the respective risk functions of the two sequences of decision rules are close to each other uniformly in θ and uniformly over all loss functions with a bounded L_∞ norm.

Brown and Low (1996) shows asymptotic equivalence of nonparametric regression and white noise. Theorem 4.1 in Brown and Low (1996) shows that in this setting, the equivalence is constructive. More specifically, Corollary 4.1 shows how to construct a sequence of procedures in nonparametric regression that is equivalent in risk to a given sequence of procedures in the white noise problem. In more or less the same time, Nussbaum (1996) showed asymptotic equivalence of density estimation and white noise. In later work on asymptotic equivalence, Brown and Zhao (2003) show equivalence of the infinite dimensional location estimation problem and nonparametric regression, and Brown et al. (2004) show equivalence of Gaussian white noise with drift, density estimation, and a Poisson process with a variable intensity.

Brown had two papers on the foundational conditional frequentist paradigm during this decade. The first, Hwang and Brown (1991), considers a frequentist confidence set $C(x)$ for θ having unconditional coverage probability $1-\alpha$, but considers reporting an *estimated coverage probability* $[1-\alpha(x)]$ (following Kiefer, 1977), which has frequentist long run validity as long as

$$E_\theta\big[1-\alpha(X)\big] \geq 1-\alpha \quad \text{(and are ideally equal).}$$

The very nice suggestion in this paper is to set up another decision problem to evaluate the "conditional accuracy" of $[1-\alpha(x)]$, by considering how close $[1-\alpha(x)]$ is to the oracle $I(\theta \supseteq C(x))$, using a loss function such as

$$L\big([1-\alpha(x)],\theta\big) = \big([1-\alpha(x)] - I(\theta \supseteq C(x))\big)^2.$$

This has become the standard method of evaluating estimated coverage probabilities.

The final Brown paper on conditional frequentist inference was Berger, Brown and Wolpert (1994). Brown's many later papers on post model selection inference could be classified as conditional frequentist inference papers, but the motivation was quite different; it had become common place to simply condition on the selected model and then perform frequentist inference, but these later papers showed how this could potentially be very wrong.

In Berger, Brown and Wolpert (1994), the basic problem of testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, based on data x from density $f(x \mid \theta)$, was considered. This had long been perceived as a problem in which frequentists and Bayesians could not agree on the answer, but the paper dramatically showed that this is not so.

Note that the usual likelihood ratio (also the Bayes factor) of H_0 to H_1 is

$$B(x) = \frac{f(x \mid \theta_0)}{f(x \mid \theta_1)}.$$

The conditional frequentist testing paradigm that was implemented chose a conditioning statistic $S = \max\{p_0, p_1\}$, where p_i is the p-value from testing H_i against the other hypothesis; the motivation is that p-values are acknowledged as providing a measure of "strength of evidence" against a hypothesis, and S is thus conditioning on sets of the same strength of evidence in the rejection and acceptance regions.

The resulting conditional frequentist test (compute Type I and Type II frequentist error probabilities, conditional on S) is shown to be given by

$$T^C = \begin{cases} \text{if } B(\mathbf{x}) \leq c, \\ \quad \text{reject } H_0 \text{ and report Type I CEP} \\ \quad \alpha(\mathbf{x}) = B(\mathbf{x})/\big(1+B(\mathbf{x})\big); \\ \text{if } B(\mathbf{x}) > c, \\ \quad \text{accept } H_0 \text{ and report Type II CEP} \\ \quad \beta(\mathbf{x}) = 1/\big(1+B(\mathbf{x})\big), \end{cases}$$

where c is the critical value at which the two p-values are equal.

These conditional error probabilities, $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$, are highly data-dependent, in that the conditioning is on S, which is typically just a two-point set, the smallest conditioning set that a frequentist could possibly use.

But the question remained as to whether this extent of conditioning was reasonable. The greatest surprise in the paper was that $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are exactly equal to the objective Bayesian posterior probabilities of H_0 and H_1, respectively, assuming that the hypotheses are given equal prior probabilities of $1/2$. Having conditional frequentists and objective Bayesians report the same error probabilities is, in some sense, the holy grail of (objectivist) statistics and doing so in a problem—in which it was thought that Bayesians and frequentists could not reconcile—was a major breakthrough in foundational (and methodological) statistics.

It is interesting that Brown did not further explicitly work on the frequentist conditional viewpoint. Perhaps he was satisfied with this major discovery that had it all.

8. NEW CENTURY: EDGEWORTH EXPANSIONS, INFERENCE AFTER MODEL SELECTION, THE ICM ADDRESS

In a sequence of articles around 2000, Brown, with coauthors, made a comprehensive examination of the coverage probability as well as the expected width of the widely used Wald confidence interval for the mean parameter in discrete and continuous one parameter exponential families, specifically the important binomial p case. It was well recognized that the Wald interval for p has poor coverage near $p = 0$ (or 1), and also even for moderate p if n was small (see, e.g., Agresti and Coull, 1998, Blyth and Still, 1983, Ghosh, 1979). Brown, Cai and DasGupta (2001, 2002), show that the coverage properties of the Wald interval are far more poor and erratic than anyone understood, even for p near 0.5 and for n as large as 200. They show, by a two term asymptotic expansion of the coverage probability as well as the expected width, that the score interval is considerably superior to the Wald interval, and so are the Jeffreys prior interval and the Agresti-Coull interval. For example, Brown, Cai and DasGupta (2002) show that the coverage probabilities $\gamma_s(n, p)$ and $\gamma_W(n, p)$ of the score and the Wald confidence interval satisfy

$$\gamma_s(n, p) - \gamma_W(n, p) = \frac{1}{4np(1-p)}\left[z^3 + \frac{z^5(1-2p)^2}{3}\right] + \text{osc.}\left(\frac{1}{\sqrt{n}}\right),$$

where z is the standard normal percentile at level $1-\alpha$, and osc.$(\frac{1}{\sqrt{n}})$ is an $O(\frac{1}{\sqrt{n}})$ oscillatory term arising from the discreteness of binomial distributions. The oscillatory term is inevitable for all nonrandomized intervals in this problem. Looking at the continuous $\frac{1}{n}$ term, we see that the coefficient is positive. This gives a remarkable theoretical explanation for why in numerical work one sees the score interval to have better coverage probabilities than the Wald interval.

There is similarly a preference order between the score and the Wald interval if one compares conditional coverage probabilities. Suppose $\beta_s(n, p)$ and $\beta_W(n, p)$ are respectively the conditional probability that the score interval would cover p when the Wald interval does not, and the conditional probability that the Wald interval would cover p when the score interval does not. Then, assuming without loss of generality that $p \geq \frac{1}{2}$, one has

$$\beta_s(n, p) - \beta_W(n, p) = \frac{1}{np(1-p)} \frac{z^3(3 + (2p-1)^2 z^2))\phi(z)}{12(1 - \Phi(z))} + \text{osc.},$$

where, again, osc. is an $O(\frac{1}{\sqrt{n}})$ oscillatory term. Notice that the coefficient of the continuous $\frac{1}{n}$ term is positive, suggesting that the score interval is preferable from this perspective too.

If one takes into account parsimony of the interval as measured by width, and ease of computation, then they recommend that the score interval be used, and use of the Wald interval for p be stopped. Brown, Cai and DasGupta (2003) unify the results for the one parameter exponential family with a quadratic variance function (Morris, 1982). These sequence of papers give an example where asymptotic expansion explains empirical and anecdotal evidence with remarkable accuracy for even moderate n. These results of Brown have been widely reported and accepted by leading texts; see, for example, Bickel and Doksum (2016), and Lehmann and Romano (2008).

A practically important issue also of foundational interest is studied in Berk, Brown and Zhao (2009) and Berk et al. (2013). In many methodological studies, the *true* model generating the data is not known when the data arrive. Exploratory analysis or more formal model selection procedures are used to select a model, and at that point the selected model is treated as the true model and susbsequent analysis is performed. For example, in an original regression model, certain coefficients may be dropped after use of the AIC or the BIC to select one model from a set of linear models, and the selected submodel is treated as if it was the true model to begin with. However, inference after model selection changes the operating characteristics of the statistical procedures and their sampling distributions. The sampling distributions become very messy, and often, or typically, cannot be written down (Pötscher, 1991, Leeb and Pötscher, 2008). Even the meaning of the post model selection parameters is ambiguous. Berk, Brown and Zhao (2009) and Berk et al. (2013) serve as extremely readable expositions to this very interesting and in some sense new area of inference and foundations. The latter article also gives concrete post model selection procedures in some concrete regression problems that are frequentistly valid.

Brown (1971) was universally considered as a deep and unimagined connection between two seemingly unrelated problems. In Brown (2002), the paper that Brown read as his invited address at the 24th ICM (International Congress of Mathematicians, Beijing, 2002), another equally deep and beautiful connection was presented. This time the connection made was between his asymptotic equivalence results and the famous Hungarian CLT which gave a strong approximation with an error bound for the sequence of normalized empirical processes for an i.i.d. sequence in one dimension. Here is a more formal description of the Hungarian CLT, often referred to as the *KMT coupling* (Komlös, Major and Tusnädy, 1975, 1976). It is useful to recall the KMT coupling result. Suppose $X_1, X_2, \dots$ is an i.i.d. sequence of real valued random variables with the CDF $F(x)$. For given $n \geq 1$, let $F_n(x) = \frac{1}{n}\sum_{i=1}^n I_{X_i \leq x}$. Define the process $\alpha_n(t) = \sqrt{n}(F_n(t) - F(t))$, $-\infty < t < \infty$. Let $B(t)$ denote a Brownian bridge on [0, 1]. Then, on a suitable common probability space, one can define $\tilde{\alpha}_n(t)$, $n \geq 1$, and a sequence of Brownian bridges $B_n(t)$, $n \geq 1$, such that:

(a)

$$\{\tilde{\alpha}_n(t), n \geq 1\} \stackrel{\mathcal{L}}{=} \{\alpha_n(t), n \geq 1\}$$

(equality in law), and for suitable C_1, C_2, λ,

(b)

$$P\Big(\sup_{-\infty<t<\infty} |\tilde{\alpha}_n(t) - B_n(F(t))| > n^{-1/2}(C_1 \log n + z)\Big) \leq C_2 e^{-\lambda z},$$

for any n and any z,

(c)

$$\sup_{-\infty<t<\infty} |\tilde{\alpha}_n(t) - B_n(F(t))| = O\left(\frac{\log n}{\sqrt{n}}\right), \quad \text{a.s.}$$

The ICM paper (Brown, 2002) describes how a construction of the KMT coupling can be done by using an explicit construction of the mappings that establish the asymptotic equivalence of two properly chosen sequences of experiments.

9. VIEWPOINTS ON STATISTICS AND ITS FUTURE

The research of Brown was influenced by the times. He worked on traditional finite sample optimality and decision theory when it was a leading doctrine of statistics. In the late eighties, Brown started to move away from admissibility, at least in his published research. Much of the most important work in the early nineties has a pronounced Bayes or minimax flavor, often in an interrelated way. Brown (1994, 1998) are lively expository articles on the conceptual role that minimaxity has played in the development and evaluation of statistical procedures. In the mid nineties, as problems with many parameters and sequence models start to emerge as the important problems, we see Brown's research move toward oracle inequalities (see, e.g., Donoho et al., 1996), asymptotic or rate minimaxity, and asymptotic equivalence. Even later, he starts to work on fully infinite dimensional problems. Brown's research evolved synergistically with decision theory.

Although decision theory formed the continuing basis of Brown's research, he was also intimately connected with other foundations of statistics. His continuing efforts in pushing the boundaries of the conditional frequentist paradigm is perhaps the best illustration. He solved many problems in this paradigm but his curiosity drove him to discover intensely deep issues in the paradigm that it may take generations to solve.

Although Brown was obviously primarily a theoretical researcher, he was also an artful applied statistician. He did well-known applied work, often Bayesian or empirical Bayesian, on sports predictions and queueing problems; he was one of the principal statisticians in the 2000 US census. The article of Linda Zhao in this volume details his applied work.

Brown was sympathetic to applications and the need for applications of statistical research and statistical theory. He did not work actively on the bootstrap, but was knowledgable about it. He did serious work on multiple comparisons and multiple testing. He did much work on financial time series models. He knew and valued the utility of simulations. In *An essay on statistical decision theory* (Brown, 2000), he says: "These simulation results can provide important practical validation of an asymptotic result or of a persuasive heuristic model. However, they do not have the intellectual force of a mathematical proof. That is, in a complex situation, I may be able to *convince* you with simulations that procedure A is better than procedure B, but rarely, if ever, can I *prove* it that way. ... Hence, the decision-theoretic challenge of finding a methodology for converting the simulational power of the computer into a tool able to deliver the persuasive force of a mathematical proof."

Brown believed that in spite of the clear and obvious trend toward stripping mathematics away from statistics, statistics will remain a fundamentally mathematical subject. He was always an optimist. In *A Conversation with Larry Brown* (DasGupta, 2005), he says: "Statistics has been and will remain useful, if anything, in more contexts than ever before. I do see a red flag on the horizon within the discipline. There seems to be a danger of fragmentation. Branches of statistics, ..., could become essentially independent subjects without a link through the fundamental core to other fragments of the field. ... I hope that statistics becomes useful in more and more areas with enough commonality that we still exist as a discipline with a unifying core."

ACKNOWLEDGMENTS

We are extremely grateful to the anonymous reviewers and the Editor for their very careful reading of our first manuscript and for their thoughtful suggestions on improving it. We are also very very grateful to Agniva Chowdhury for helping us very significantly with final formatting of this article.

Research partially supported by NSF Grant DMS-1407775 and ELS-PU 90014395.

REFERENCES

AGRESTI, A. and COULL, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *Amer. Statist.* **52** 119–126. MR1628435

BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Stat.* **25** 423–462. MR0063630

BARANKIN, E. W. and MAITRA, A. P. (1963). Generalization of the Fisher–Darmois–Koopman–Pitman theorem on sufficient statistics. *Sankhya, Ser. A* **25** 217–244. MR0171342

BARRON, A. R. (1986). Entropy and the central limit theorem. *Ann. Probab.* **14** 336–342. MR0815975

BERGER, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226. MR0397940

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR0804611

BERGER, J. O., BROWN, L. D. and WOLPERT, R. L. (1994). A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing. *Ann. Statist.* **22** 1787–1807. MR1329168

BERGER, J. O. and STRAWDERMAN, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Ann. Statist.* **24** 931–951. MR1401831

BERGER, J., BOCK, M. E., BROWN, L. D., CASELLA, G. and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771. MR0443156

BERK, R., BROWN, L. D. and ZHAO, L. (2009). Statistical inference after model selection. *J. Quant. Criminol.* **26** 217–236.

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309. MR0630112

BICKEL, P. J. and DOKSUM, K. A. (2016). *Mathematical Statistics—Basic Ideas and Selected Topics. Vol.* 2, 2nd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3287337

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Stat.* **22** 393–399. MR0043418

BLYTH, C. R. and STILL, H. A. (1983). Binomial confidence intervals. *J. Amer. Statist. Assoc.* **78** 108–116. MR0696854

BOROVKOV, A. A. and SAKHANENKO, A. I. (1980). Estimates for averaged quadratic risk. *Probab. Math. Statist.* **1** 185–195. MR0626310

BROWN, L. (1964). Sufficient statistics in the case of independent random variables. *Ann. Inst. Statist. Math.* **35** 1456–1474. MR0216611

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Inst. Statist. Math.* **37** 1087–1136. MR0216647

BROWN, L. (1967). The conditional level of Student's t test. *Ann. Inst. Statist. Math.* **38** 1068–1071. MR0214210

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Inst. Statist. Math.* **42** 855–903. MR0286209

BROWN, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Ann. Statist.* **6** 59–71. MR0471160

BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960–994. MR0536501

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300. MR0630111

BROWN, L. D. (1982). A proof of the central limit theorem motivated by the Cramér–Rao inequality. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 141–148. North-Holland, Amsterdam. MR0659464

BROWN, L. D. (1983). Comments on "The Robust Bayesian Viewpoint" by Berger, J. O.. In *Robustness of Bayesian Analyses* (J. B. Kadane, ed.) 126–133. Elsevier, Amsterdam.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* (S. S. Gupta, ed.) **9**. IMS, Hayward, CA. MR0882001

BROWN, L. D. (1988). The differential inequality of a statistical estimation problem. In *Statistical Decision Theory and Related Topics, IV, Vol.* 1 (*West Lafayette, Ind.*, 1986) (S. S. Gupta and J. O. Berger, eds.) 299–324. Springer, New York. MR0927109

BROWN, L. D. (1990). The 1985 Wald Memorial Lectures. An ancillarity paradox which appears in multiple linear regression. *Ann. Statist.* **18** 471–538. MR1056325

BROWN, L. D. (1994). Minimaxity, more or less. In *Statistical Decision Theory and Related Topics, V* (*West Lafayette, IN*, 1992) (S. S. Gupta and J. O. Berger, eds.) 1–18. Springer, New York. MR1286291

BROWN, L. D. (1998). *Minimax Theory, in Encyclopedia of Biostatistics* (P. Armitage and T. Colton, eds.). Wiley, New York.

BROWN, L. D. (2000). An essay on statistical decision theory. *J. Amer. Statist. Assoc.* **95** 1277–1281. MR1825275

BROWN, L. D., ed. (2002). An analogy between statistical equivalence and stochastic strong limit theorems. Invited address at the International Congress of Mathematicians, Beijing, August, 2002.

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133. MR1861069

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30** 160–201. MR1892660

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2003). Interval estimation in exponential families. *Statist. Sinica* **13** 19–49. MR1963918

BROWN, L. D. and COHEN, A. (1981). Inadmissibility of large classes of sequential tests. *Ann. Statist.* **9** 1239–1247. MR0630106

BROWN, L. D., COHEN, A. and SAMUEL-CAHN, E. (1983). A sharp necessary condition for admissibility of sequential tests—necessary and sufficient conditions for admissibility of SPRTs. *Ann. Statist.* **11** 640–653. MR0696075

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1979). Monotonicity of Bayes sequential tests. *Ann. Statist.* **7** 1222–1230. MR0550145

BROWN, L. D. and FARRELL, R. H. (1990). A lower bound for the risk in estimating the value of a probability density. *J. Amer. Statist. Assoc.* **85** 1147–1153. MR1134513

BROWN, L. D. and GAJEK, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** 1578–1594. MR1074424

BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics, III*, *Vol*. 1 (*West Lafayette*, *Ind*., 1981) 205–230. Academic Press, New York. MR0705290

BROWN, L. D. and LIU, R. C. (1993). Bounds on the Bayes and minimax risk for signal parameter estimation. *IEEE Trans*. *Inform*. *Theory* **39** 1386–1394.

BROWN, L. D. and LOW, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann*. *Statist*. **19** 329–337. MR1091854

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann*. *Statist*. **24** 2384–2398. MR1425958

BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. *Ann*. *Statist*. **1** 902–912. MR0432846

BROWN, L. D. and ZHAO, L. (2003). Direct asymptotic equivalence of nonparametric regression and the infinite-dimensional location problem. Preprint Univ. Pennsylvania, Philadelphia, PA.

BROWN, L. D., CARTER, A. V., LOW, M. G. and ZHANG, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann*. *Statist*. **32** 2074–2097. MR2102503

BROWN, L., DASGUPTA, A., HAFF, L. R. and STRAWDERMAN, W. E. (2006). The heat equation and Stein's identity: Connections, applications. *J*. *Statist*. *Plann*. *Inference* **136** 2254–2278. MR2235057

CLEVENSON, M. L. and ZIDEK, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *J*. *Amer*. *Statist*. *Assoc*. **70** 698–705. MR0394962

COHEN, A. (1965). Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution. *Ann*. *Inst*. *Statist*. *Math*. **36** 78–87. MR0172399

DASGUPTA, A. (1983). Bayes and minimax estimation in one and multiparameter families. Ph.D. thesis, Indian Statistical Institute.

DASGUPTA, A. (2005). A conversation with Larry Brown. *Statist*. *Sci*. **20** 193–203. MR2183449

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. *Springer Texts in Statistics*. Springer, New York. MR2664452

DASGUPTA, A. and SINHA, B. K. (1986). Estimation in the multiparameter exponential family: Admissibility and inadmissibility results. *Statist*. *Decisions* **4** 101–130. MR0848322

DIACONIS, P. and STEIN, C. (1982). Lecture notes on statistical decision theory. Dept. Statistics, Stanford Univ., Stanford, CA.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann*. *Statist*. **18** 1416–1437. MR1062717

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann*. *Statist*. **24** 508–539. MR1394974

DYNKIN, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions. In *Select*. *Transl*. *Math*. *Statist*. *and Probability*, *Vol*. 1 17–40. IMS and Amer. Math. Soc., Providence, RI. MR0116407

EATON, M. L. (1992). A statistical diptych: Admissible inferences—recurrence of symmetric Markov chains. *Ann*. *Statist*. **20** 1147–1179. MR1186245

FISHER, R. A. (1923). On the mathematical foundations of theoretical statistics. *Philos*. *Trans*. *R*. *Soc*. *Lond*. *Ser*. *A* **222** 309–368.

GHOSH, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *J*. *Amer*. *Statist*. *Assoc*. **74** 894–900. MR0556485

HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann*. *Math*. *Stat*. **20** 225–241. MR0030730

HSUAN, F. C. (1979). A stepwise Bayesian procedure. *Ann*. *Statist*. **7** 860–868. MR0532249

HWANG, J. T. and BROWN, L. D. (1991). Estimated confidence under the validity constraint. *Ann*. *Statist*. **19** 1964–1977. MR1135159

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc*. 4*th Berkeley Sympos*. *Math*. *Statist*. *and Prob*., *Vol*. *I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JOHNSON, B. MCK. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann*. *Inst*. *Statist*. *Math*. **42** 1579–1587. MR0418300

JOHNSON, O. (2004). *Information Theory and the Central Limit Theorem*. Imperial College Press, London. MR2109042

JOHNSTONE, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann*. *Statist*. **12** 1173–1198. MR0760682

JOHNSTONE, I. and LALLEY, S. (1984). On independent statistical decision problems and products of diffusions. *Z*. *Wahrsch*. *Verw*. *Gebiete* **68** 29–47. MR0767442

KAGAN, A. M., LINNIK, Y. V. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York. MR0346969

KARLIN, S. (1958). Admissibility for estimation with quadratic loss. *Ann*. *Inst*. *Statist*. *Math*. **29** 406–436. MR0124101

KIEFER, J. (1977). Conditional confidence statements and confidence estimators. *J*. *Amer*. *Statist*. *Assoc*. **72** 789–827. MR0518611

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z*. *Wahrsch*. *Verw*. *Gebiete* **32** 111–131. MR0375412

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z*. *Wahrsch*. *Verw*. *Gebiete* **34** 33–58. MR0402883

KOROSTELEV, A. and KOROSTELEVA, O. (2011). *Mathematical Statistics*. *Graduate Studies in Mathematics*, *Asymptotic Minimax Theory* **119**. Amer. Math. Soc., Providence, RI. MR2767163

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411

LEEB, H. and PÖTSCHER, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* **24** 338–376. MR2422862

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York. MR0107933

LEHMANN, E. L. and ROMANO, J. P. (2008). *Testing Statistical Hypotheses*. Springer, New York.

LEHMANN, E. L. and STEIN, C. M. (1953). The admissibility of certain invariant statistical tests involving a translation parameter. *Ann*. *Math*. *Stat*. **24** 473–479. MR0056249

LEVIT, B. Y. (1987). On bounds for the minimax risk. In *Probability Theory and Mathematical Statistics*, *Vol*. *II* (*Vilnius*, 1985) 203–216. VNU Sci. Press, Utrecht. MR0901534

LIU, R. C. and BROWN, L. D. (1993). Nonexistence of informative unbiased estimators in singular problems. *Ann. Statist.* **21** 1–13. MR1212163

MEEDEN, G., GHOSH, M., SRINIVASAN, C. and VARDEMAN, S. (1989). The admissibility of the Kaplan–Meier and other maximum likelihood estimators in the presence of censoring. *Ann. Statist.* **17** 1509–1531. MR1026297

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10** 65–80. MR0642719

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959

OLKIN, I. and SOBEL, M. (1979). Admissible and minimax estimation for the multinomial distribution and for k independent binomial distributions. *Ann. Statist.* **7** 284–290. MR0520240

PÖTSCHER, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7** 163–185. MR1128410

RUKHIN, A. L. (1995). Admissibility: Survey of a concept in progress. *Int. Stat. Rev.* **63** 95–115.

SINHA, B. K. and GUPTA, A. D. (1984). Admissibility of generalized Bayes and Pitman estimates in the nonregular family. *Comm. Statist. Theory Methods* **13** 1709–1721. MR0742524

SKIBINSKY, M. and RUKHIN, A. L. (1989). Admissible estimators of binomial probability and the inverse Bayes rule map. *Ann. Inst. Statist. Math.* **41** 699–716. MR1039400

SOBEL, M. (1953). An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Stat.* **24** 319–337. MR0056899

STEIN, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Inst. Statist. Math.* **30** 970–979. MR0109392

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

TSYBAKOV, A. B. (2004). *Introduction à L'estimation Non-paramétrique. Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **41**. Springer, Berlin. MR2013911

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275