An Essay on Statistical Decision Theory

Lawrence D. BROWN

1. A 1966 QUOTATION

The middle third of this century marks the summit of research in statistical decision theory. Consider this 1966 quotation from the foreword to the volume Early Statistical Papers of J. Neyman (Neyman 1967), signed "students of J. N. at Berkeley":

> The concepts of confidence intervals and of the Neyman–Pearson theory have proved immensely fruitful. A natural but far reaching extension of their scope can be found in Abraham Wald's theory of statistical decision functions. The elaboration and application of the statistical tools related to these ideas has already occupied a generation of statisticians. *It continues to be the main lifestream of theoretical statistics* [italics mine].

The italicized assertion is the focus of this vignette. Is the assertion still true today? If not, what is the current status of statistical decision theory, and what position is it likely to hold in the coming decades? Any attempt to answer these questions requires a definition of "statistical decision theory." Indeed, the answers will be largely driven by how broadly—or narrowly—the boundaries of decision theory are drawn.

2. THE SCOPE OF STATISTICAL DECISION THEORY

The term "statistical decision theory" appears to be a condensation of Wald's phrase "the theory of statistical decision functions," which occurs, for example, in the preface to his monograph (Wald 1950) as well as earlier in Wald (1942). Wald viewed his "theory" as a codification and generalization of the theory of tests and confidence intervals already developed by Neyman, often in collaboration with E. Pearson. For clear statements of Wald's view see the last two paragraphs of the introduction to Wald's pivotal paper (Wald 1939), and section 1.7 of his monograph (Wald 1950). The vignette on hypothesis testing by Marden presents an excellent review of the various manifestations of hypothesis testing. It is hard to choose a favorite among the wonderful Neyman–Pearson papers on the foundations of testing and confidence intervals, but I pick the works by Neyman and Pearson (1933) on testing and Neyman (1935) on confidence intervals.

The frequentist approach is the cornerstone of Neyman's statistical philosophy. Neyman (1941) provided a graphic demonstration and explanation. Thus one begins the analysis of any statistical situation with a family of possible distributions for the data. In the presence of a finite-dimensional parameterization, this can be written as $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$, but the existence of such a parameterization is only an often-useful convenience, rather than a formal necessity. One then hypothesizes a possible rule for solving the problem (a test or confidence interval, or a decision function in Wald's more general terminology). The key step is to calculate the distribution of outcomes from that rule as if the true parameter were a fixed value, θ, and compare the results of such a calculation at various θ for various possible decision rules.

That "as if θ were fixed" qualification caused considerable confusion in the early years of the theory, and often continues to do so. It does not mean that θ *is* fixed. Neyman and Pearson frequently returned to emphasize that this type of calculation would guarantee validity "irrespective of the a priori truth." They denied any presumption that the statistician would be faced with a long sequence of independent repetitions of the situation, all having the same value of θ. For example, the landmark Clopper and Pearson (1934) article on confidence intervals for a binomial parameter discusses as a particular example what happens if the Clopper–Pearson prescription is used in a situation where the unknown parameter, p, can take the values $1/3$, $1/2$, and $2/3$ with a hypothesized skewed a priori distribution. The point of this calculation is to vividly demonstrate with an example the claim that the proposed intervals have coverage probability at least the nominal value, and this fact "does not depend on any a priori knowledge regarding possible values of p." A subsidiary goal may have been to investigate by how much this nominal value would be exceeded in such a plausible example. (The answer was that the true coverage was .9676, as opposed to the nominal value of .95.) A later discussion related to this general issue appears in Neyman (1952, p. 211). Brown, Cai, and Das Gupta (1999a,b) recently reexamined the problem of binomial confidence intervals and included a reassessment of the Clopper–Pearson proposal.

3. THE DECISION THEORETIC SPIRIT

According to the foregoing, the spirit of decision theory is pervasive in contemporary statistical research. Common manifestations include both mathematical and numerical attempts to check the frequentist performance of proposed procedures. This includes comparative investigations of level and power for hypothesis tests or of precision of proposed estimators as, for example, might occur in a Monte Carlo comparison of variances and biases. In particular, any presentation of statistical tests that mentions power is an embodiment of this spirit.

The frequentist interpretation of confidence intervals (and sets) relies on Neyman's previously cited articles as well as Wilson (1927) and the previously cited work of Clopper and Pearson. Note also the general formulation by Wald and Wolfowitz (1939, 1941) of nonparametric confidence

Lawrence D. Brown is Miers Busch Professor of Statistics, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104-6302 (E-mail: *lbrown@stat.wharton.upenn.edu*).

Journal of the American Statistical Association
December 2000, Vol. 95, No. 452, Vignettes

bands for a cumulative distribution function, and their discovery that a fundamental probabilistic question they could not solve had earlier been settled by Kolmogorov (1933) and Smirnov (1939).

4. SEQUENTIAL ANALYSIS

Wald was justifiably proud of his formulation of sequential decision problems and his solution of fundamental issues there, as in the optimal property of the sequential probability ratio test presented by Wald and Wolfowitz (1950). Wald (1947) cited a few historical precedents in the work of others, but there is no question that this entire statistical area was his creation. The spirit of his development survives in parts of contemporary statistics and even flourishes in some, as in the methodology of early stopping for (sequential) clinical trials; for example, via group sequential testing or stochastic curtailment. Jennison and Turnbull (1990) contains a fairly recent review of clinical trials, and Siegmund (1994) provides a broader review of the current status of Wald's sequential analysis. As a prelude to my later discussion, let me already note that in this area the analytic tools originally developed by Wald (and later extended and refined by others) survive as essential building blocks in contemporary research.

5. MINIMAXITY AS A THEME AND BENCHMARK

I have argued (Brown 1994) that the minimax idea has been an essential foundation for advances in many areas of statistical research. These include general asymptotic theory and methodology, hierarchical models, robust estimation, optimal design, and nonparametric function analysis. The vignette on minimaxity by W. Strawderman clarifies and reinforces this assertion. Hence here I do not specifically pursue this crucial manifestation of the spirit of decision theory, although it is very much present in the nonparametric examples I describe.

6. BAYESIAN STATISTICS

Not so many years ago, "bayesian statistics" was frequently viewed as the antithesis of "frequentist statistics," and the feeling among some was that their favorite of these two theories would eventually triumph as the other failed ignominiously. It now is apparent that this will not happen. There is much evidence that we are currently in the midst of a productive amalgamation of these two schools of statistics.

The vignette on Bayesian analysis by Berger describes several different approaches to Bayesian analysis. Interestingly, none of these is directly the pure frequentist approach, in which the prior is a given distribution with the same frequentist validity as the family of distributions, $\mathcal{F}$. Such a situation is a conceptual and sometimes realistic possibility, but modern Bayesian statistical techniques are intended to apply far beyond this possible scenario. Lehmann and Casella (1998, p. 226) provided for further discussion of this state of affairs.

According to the discussion that began this essay, a pure frequentist Bayesian approach, where realistic, is firmly within the decision-theoretic realm. Other fundamental Bayesian approaches are not, even if they may involve loss functions and even though they may be justified, at least in part, by the decision-theoretic version of Bayes theorem that says the Bayes procedure minimizes the expected risk. However, most of these approaches are neutral rather than antithetical with respect to decision theory. Many of them involve the use of "objective" priors, such as the Jeffreys prior or the Bernardo (1979) reference priors. These approaches customarily generate decisions. As such, they can be viewed as powerful and heuristically appealing *mechanisms* for generating plausible decision rules. Having used such a device, the question remains as to how well the rule that it generates will actually perform in a suitable range of situations like the one at hand. This decision-theoretic question is implicit but equally pertinent in varieties of robust Bayesian analyses and is explicit in the realm of Γ-minimax procedures (see Hodges and Lehmann 1952 for an early decision-theoretic formulation closely related to the Γ-minimax idea). Answering such a question is increasingly (and properly) seen as requiring frequentist-style investigations of comparative risk through simulations or theoretical calculations. Hence what is emerging is a figurative marriage of Bayes and Wald.

7. THE DECISION THEORETIC TOOLKIT

I have been arguing in general terms that Neyman and Wald (and also their collaborators and immediate students) had a particular perspective on statistics, and that this perspective is alive and thriving in contemporary statistics. Nevertheless, many statisticians apparently feel that statistical decision theory is moribund—or even already dead.

Through the decades, various technical analytical tools have been developed by avowedly decision-theoretic researchers. An assertion that decision theory is dying is probably more focused on this toolkit than it is on the decision-theoretic spirit I discussed earlier. Even here, the assertion seems to me in the main to be drastically mistaken, though one might point to certain tools that are far less broadly useful than might have been expected at the time they were being developed. I have in mind as an example of the latter the general complete-class theory involving characterization of admissible procedures. (Even here there are significant recent contributions relevant to contemporary methodology; see, e.g., Berger and Strawderman 1996 and Zhao 1999.)

This "toolkit" contains a vast array of analytical weapons for a variety of situations. Furthermore, it has continually extended and expanded from its original extent and form. It is impossible in the format of this short, broad survey to carefully trace the development of even a single important tool from its origins with Neyman or Wald (and often before them as well). So I discuss only some examples from one particular area of research, to try to emphasize the vitality and importance of that legacy.

8. NONPARAMETRIC FUNCTION ANALYSIS

8.1 Rates of Convergence

Nonparametric function analysis includes such topics as nonparametric regression, density estimation, image reconstruction, and aspects of pattern recognition. As a manifestation of the nonparametric/robust paradigm, it is already heavily infused with the influence of decision theory. For confirmation, note that on Efron's (1998) "barycentric picture of modern statistical research," the topic "robustness, nonparametrics" describes the point most heavily weighted on the "frequentist" axis, as opposed to the "Bayesian" and "Fisherian" axes. There should be no debate on this point. But I want to go further, and describe how several items in the basic toolkit are of use here.

First, the Waldian notion of loss and risk pervades the topic. A fundamental feature of this area is the presence of rates of convergence slower than the usual parametric standard of $1/\sqrt{n}$. These rates of convergence are for the *risk* under any of various loss functions. In particular, the asymptotics cannot be formulated in Fisherian terms of efficiency, because the optimal-rate risks under (integrated) quadratic loss must balance squared bias as well as variance. Nevertheless, Fisher information and the Cramer–Rao inequality occasionally can be useful (as in Brown and Low 1991), but the spirit and content of such a treatment are more directly descended from the admissibility argument of Hodges and Lehmann (1951) and the two-dimensional admissibility argument of Stein (1956b).

A precise description of these convergence rates nearly demands a minimax formulation. An exception is the early formulation of Rosenblatt (1956) and Parzen (1962), looking at optimal rates available from the use of nonnegative kernels. Brown, Low, and Zhao (1997) tried to explain why this is more emphatically so here than in the classical parametric theory.

Despite its success, there is a not-untypical shortcoming in a literal adoption of this minimax formulation—it may be unhealthily conservative. The formulation assumes that the unknown function (e.g., regression function or density) belongs to a suitably bounded class, usually explicitly involving some sort of smoothness restriction. For a basic example, in a one-dimensional setting, the assumption might be that

$$\int (f''(x))^2\, dx \le B$$

for a prespecified, but possibly large value of B. Corresponding to this class is an asymptotically minimax value and procedure. An adequate practical approach to this asymptotic ideal is achievable with computationally feasible procedures of various types, such as kernels, splines, and orthogonal series estimators (including wavelets). However, a procedure that is minimax in this sense, or close to it, may perform relatively poorly in practice. To see why, consider a typical simple task such as estimating an unknown probability density. One may suspect that the true density is quite "smooth"—perhaps it is thought to have a shape similar to a simple mixture of normal distributions. Various standard optimal minimax convergence rate procedures do not utilize that suspicion. Instead, they protect against the possibility that the density is as extremely wiggly as allowed by the foregoing smoothness restriction. Such extremely wiggly densities may be felt to be a priori highly unlikely and/or not very interesting. A current tidal wave of research into "adaptive" estimators is an attempt to create procedures that circumvent this shortcoming by being simultaneously nearly minimax for a broad inventory of smoothness classes.

8.2 Hardest Linear Subproblem

The minimax value just mentioned can be discovered to within a startling degree of accuracy through very beautiful application of a fundamental device due to Wald in a form proposed by Stein (1956a) (see Donoho 1994; Donoho and Liu 1991; Donoho, Liu, and MacGibbon 1990). Consider the entirely classical parametric situation of an observation from a multivariate normal distribution with identity covariance matrix but unknown mean. Suppose that one desires to estimate some linear functional of that mean—for example, its first coordinate—and uses ordinary quadratic loss. Further, suppose the mean is known to lie in a bounded convex set, S, symmetric about the origin.

Temporarily restrict the class of available estimators to be linear. Wald's standard tool for discovering minimax values is to establish a least favorable distribution. Here that least favorable distribution turns out to be supported on a one-dimensional subset of S, say H. That is, for linear estimators, there is a "hardest one-dimensional subproblem." The minimax risk for the problem on S unrestricted to linear estimators cannot be less than that for the unrestricted problem on H. One then gets to look at an even more classical problem. Let $X \sim \mathrm{N}(\theta, 1)$ with $\theta \in (-a, a)$. What is the minimax risk (under squared error loss), and what is its relation to the minimax risk among linear procedures? This question had been studied by Casella and Strawderman (1981), among others. Extending their result, Donoho et al. (1990) showed that the ratio of the two risks is never greater than 1.25.

This very classical (but nevertheless recent) minimax theorem about estimation of a multivariate normal mean can then be carried into the nonparametric realm with the aid of a further set of decision-theoretic tools for asymptotics largely developed by Le Cam (see, e.g., LeCam 1953, 1986).

The end result from using this decision theoretic arsenal is a powerful result enabling one to calculate the (asymptotic) minimax risk to within a factor of 1.25 and, even more important, to know how to get a (linear) procedure that comes within this factor of being minimax. Interestingly, this idea works for various loss functions in addition to squared error (see Donoho 1994).

8.3 Asymptotic Equivalence

Further utilization of Le Cam's decision theoretic ideas has proved valuable in this area. LeCam (1953, 1964, 1986) created a general concept of equivalence of statistical problems within the framework of Wald's decision theory. With the aid of additional inequalities also due at least partly to Le Cam, one can unify and greatly simplify asymptotic investigations similar to those sketched earlier.

To explain, I return briefly to a more classical parametric setting. Asymptotic questions in such settings very often can be efficiently reduced to appropriate questions involving only normal distributions. These normal distribution settings can effectively be viewed as the limit of the original problem as the sample size tends to infinity. This aspect of statistical theory has its roots well before Neyman and Wald and, I believe, lies well outside the scope of what should be considered characteristically decision theoretic. Nevertheless, both Neyman and Wald recognized its importance, made important contributions in the area, and incorporated asymptotical analyses into their basic theories. Prominent examples are Wald's (1943) proof of the asymptotic optimality of the likelihood ratio test and Neyman's (1949) explication of "best asymptotically normal" procedures.

In an analogous fashion the stochastic formulation of white noise with drift can serve as a useful, unified limit for nonparametric formulations like those already mentioned. Research on this topic is ongoing, but Brown and Low (1996) and Nussbaum (1996) have given basic results.

By all measures, wavelets provide a powerful new tool for nonparametric function analysis. The continuing development of this tool has evolved out of a combination of the function-analytic topic of wavelet bases with an intensive and extensive dose of decision theory, along with a careful evaluation of the practical problems requiring the statistical techniques being created. The debt to decision theory will be immediately evident to any reader of the fundamental works of Donoho and Johnstone and collaborators (e.g., Donoho and Johnstone 1995, 1998).

In this development there is a continuing expansion of the ideas present in the minimax treatment of Donoho et al. described previously. I will not go into further detail about that. Instead, I mention two quite different decision-theoretic aspects. One is the incorporation of Stein's unbiased estimate of the risk. Stein's powerful technical tool was developed as a more effective way of handling certain admissibility questions that arose following his surprising discovery of the inadmissibility of the usual estimator of a multivariate normal mean (see Brown 1979; James and Stein 1961; Stein 1956, 1973, 1981). Donoho and Johnstone (1995) cited this technique as the defining element of a class of wavelet estimators that they called SURE-type estimators.

A more unexpected combination of techniques is the application of Benjamini and Hochberg (1995) in the construction of wavelet estimators. Benjamini and Hochberg created an interesting new proposal for the problem of multiple comparisons. Their idea is to control what they term the "false discovery rate" (FDR). The issue of multiple comparisons is heavily decision theoretic in orientation and development. Nevertheless, the basic problem that Benjamini and Hochberg addressed would seem quite distant from the issues relevant to wavelet estimation. But Abramowich and Benjamini (1995), Abramowich, Benjamini, Donoho, and Johnstone (1999), and Johnstone (1999) described an important and productive connection between FDR and adaptive wavelet estimation.

9. MONTE CARLO INVESTIGATIONS: A CHALLENGE FOR FUTURE THEORY

I have been focusing on what I termed the decision-theoretic toolkit. I believe that this kit is largely complete and that the focus of future research will involve use of these tools, as in the foregoing stories about nonparametric functional analysis. Of course, I could be wrong. There is some interesting, continuing toolbuilding going on (as, e.g., in Eaton 1992). Then, too, maybe some brand new tool is just waiting to be discovered—perhaps a biased estimate of risk that will prove even more useful than Stein's unbiased estimate. Or maybe an accumulation of results developing out of the current toolkit will become recognized as an independently powerful tool on its own.

However, there is at least one place where I think we are lacking a general decision-theoretic tool of a new sort. As I have noted, risk comparisons of proposed procedures are often performed via simulation. (We may be simulating the performance of the respective proposals in terms of level and power or expected coverage and, sometimes, expected length or size, or in terms of bias and variance, or just to discover their sampling distribution, leaving the determination of comparative risk to the reader.)

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. This is primarily because simulations can be performed only at specific alternatives and sample sizes, and there are usually too many such alternatives of interest for an adequate simulation to be performed at each one. 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.

REFERENCES

Abramowich, F., and Benjamini, Y. (1995), "Threshholding of Wavelet Coefficients as Multiple Testing Procedures," in *Wavelets and Statistics*, ed. A. Antoniadis, New York: Springer-Verlag, pp. 5–14.

Abramowich, A., Benjamini, Y., Donoho, D., and Johnstone, I. M. (1999), "Sparsity and Threshholding via False Discovery Control," preprint.

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Ser. B, 57, 289–300.

Berger, J. O., and Strawderman, W. E. (1996), "Choice of Hierarchical Priors: Admissibility in Estimation of Normal Means," *The Annals of Statistics*, 24, 931–951.

Bernardo, J. M. (1979), "Reference Prior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society*, Ser. B, 41, 113–147.

Brown, L. D. (1979), "A Heuristic Method for Determining Admissibility of Estimators, With Applications," *The Annals of Statistics*, 7, 960–994.

——— (1994), "Minimaxity, More or Less," in *Statistical Decision Theory and Related Topics*, eds. S. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 1–18.

Brown, L. D., Cai, T. T., and DasGupta, A. (1999a), "Interval Estimation

for a Binomial Proportion," submitted for publication.
——— (1999b), "Confidence Intervals for a Binomial Proportion and Edgeworth Expansions," submitted for publication.
Brown, L. D., and Low, M. G. (1991), "Information Inequality Bounds on the Minimax Risk With an Application to Nonparametric Regression," *The Annals of Statistics*, 19, 329–337.
——— (1996), "Asymptotic Equivalence of Nonparametric Regression and White Noise," *The Annals of Statistics*, 24, 2384–2398.
Brown, L. D., Low, M. G., and Zhao, L. H. (1997), "Superefficiency in Nonparametric Function Estimation," *The Annals of Statistics*, 25, 2607–2625.
Casella, G., and Strawderman, W. E. (1981), "Estimating a Bounded Normal Mean," *The Annals of Statistics*, 9, 870–878.
Clopper, C. J., and Pearson, E. S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404–413.
Donoho, D. L. (1994), "Statistical Estimation and Optimal Recovery," *The Annals of Statistics*, 22, 238–270.
Donoho, D. L., and Johnstone, I. M. (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of the American Statistical Association*, 90, 1200–1224.
——— (1998), "Minimax Estimation via Wavelet Shrinkage," *The Annals of Statistics*, 26, 879–921.
Donoho, D. L., and Liu, R. C. (1991), "Geometrizing Rates of Convergence III," *The Annals of Statistics*, 19, 668–701.
Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990), "Minimax Risk for Hyperrectangles," *The Annals of Statistics*, 18, 1416–1437.
Eaton, M. L. (1992), "A Statistical Diptych: Admissible Inferences–Recurrence of Symmetric Markov Chains," *The Annals of Statistics*, 20, 1147–1179.
Efron, B. (1998), "R. A. Fisher in the 21st Century" (with discussion), *Statistical Science*, 13, 95–113.
Hodges, J. L., Jr., and Lehmann, E. L. (1950), "The Use of Previous Experience in Reaching Statistical Decisions," *Annals of Mathematical Statistics*, 23, 396–407.
——— (1951), "Some Applications of the Cramer–Rao Inequality," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1, Berkeley, CA: University of California Press, pp. 13–22.
James, W., and Stein, C. (1961), "Estimation With Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, Berkeley, CA: University of California Press, pp. 311–319.
Jennison, C., and Turnbull, B. W. (1990), "Statistical Approaches to Interim Monitoring of Medical Trials: A Review and Commentary," *Statistical Science*, 5, 299–317.
Johnstone, I. M. (1999), "False Discovery Rates and Adaptive Estimation," special invited paper presented at the 1999 Joint Statistical Meetings in Baltimore.
Kolmogorov, A. (1933), "Sulla Determinazione Empirica di una Leggi di Distribuzione," *Giornale dell'Instituto Italiana degli Attuari*, 4, 83–91.
Le Cam, L. (1953), *On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes Procedures*, Berkeley, CA: University of California Press, pp. 277–230.
——— (1964), "Sufficiency and Approximate Sufficiency," *Annals of Mathematical Statistics*, 35, 1419–1455.
——— (1986), *Asymptotic Methods in Statistical Decision Theory*, New York: Springer-Verlag.
Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation* (2nd ed.), New York: Springer.
Neyman, J. (1935), "On the Problem of Confidence Intervals," *Annals of Mathematical Statistics*, 6, 111–116.
——— (1941), "Fiducial Argument and the Theory of Confidence Intervals," *Biometrika*, 32, 128–150.
——— (1949), "Contribution to the Theory of the Chi-Square Test," in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 1945, Berkeley, CA: University of California Press, pp. 239–273.
——— (1952), *Lectures and Conferences on Mathematical Statistics and Probability*, Washington, D.C.: U.S. Department of Agriculture.
——— (1967), *A Selection of Early Statistical Papers of J. Neyman*, Berkeley, CA: University of California Press.
Neyman, J., and Pearson, E. S. (1933), "The Testing of Statistical Hypotheses in Relation to Probabilities a Priori," *Proceedings of the Cambridge Philosophical Society*, 29, 492–510.
Nussbaum, M. (1996), "Asymptotic Equivalence of Density Estimation and Gaussian White Noise," *The Annals of Statistics*, 24, 2399–2430.
Parzen, E. (1962), "On the Estimation of a Probability Density and Mode," *Annals of Mathematical Statistics*, 33, 1065–1076.
Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 832–837.
Siegmund, D. (1994), "A Retrospective of Wald's Sequential Analysis—Its Relation to Changepoint Detection and Sequential Clinical Trials," in *Statistical Decision Theory and Related Topics*, eds. S. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 19–33.
Smirnov, N. V. (1939), "On the Estimation of the Discrepancy Between Empirical Curves of Distributions for Two Independent Samples," *Bulletin of the University of Moscow*, 2(2), 3–14.
Stein, C. (1956a), "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, pp. 187–196.
——— (1956b), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, pp. 197–206.
——— (1973), "Estimation of the Mean of a Multivariate Distribution," in *Proceedings of the Prague Symposium on Asymptotic Statistics*, pp. 345–381.
——— (1981), "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, 9, 1135–1151.
Wald, A. (1939), "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, 10, 299–326.
——— (1942), *On the Principles of Statistical Inference*, Notre Dame, IN.
——— (1943), "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.
——— (1947), *Sequential Analysis*, New York: Wiley.
——— (1950), *Statistical Decision Functions*, New York: Wiley.
Wald, A., and Wolfowitz, J. (1939), "Confidence Limits for a Continuous Distribution Function," *Annals of Mathematical Statistics*, 10, 105–118.
——— (1941), "Note on Confidence Limits for Continuous Distribution Functions," *Annals of Mathematical Statistics*, 12, 118–119.
——— (1950), "Bayes Solution of Sequential Decision Problems," *Annals of Mathematical Statistics*, 21, 82–89.
Wilson, E. B. (1927), "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, 22, 209–212.
Zhao, L. H. (1999), "Improved Estimators in Nonparametric Regression Problems," *Journal of the American Statistical Association*, 94, 164–173.