Erich L. Lehmann's Work on Decision Theory

Javier Rojo

This chapter collects some of Erich's work on areas that attracted his attention during the early years in his professional career. Thus, the papers discuss the concepts of completeness (minimal complete families and complete sufficient statistics), minimal sufficiency, admissibility, invariance, unbiasedness, and minimaxity.

In Lehmann (1947) Erich argued that, in the absence of a uniformly most powerful (UMP) test, it is then a difficult choice to select a "best" test. Since there may be too many choices, it was proposed to reduce the totality of possible tests, by restricting attention to a class of tests $\mathcal{F}$ such that for any test ϕ of a hypothesis $\mathcal{H}$ of a given significance level, there is a corresponding test ϕ^* in $\mathcal{F}$ with power uniformly larger than or equal to the power of ϕ; and if $\phi_1, \phi_2 \in \mathcal{F}$, then the power of ϕ_1 is not uniformly larger than the power of ϕ_2 and vice versa. That is, in the absence of a UMP test, $\mathcal{F}$ is the smallest class from which to select a test when the selection is made based on power. Although there are other important ideas in the paper, this concept became a central part of Wald's statistical decision theory. In Lehmann (2008), Erich writes that

> *Wald at that time had resumed work on the general theory of statistical inference he had outlined in his 1939 paper. He found that my suggestion fit well into his general framework, and he magically transformed it into a theorem of great beauty and generality, which became one of the principal pillars of his decision theory.*

Motivated by the lack of specific examples of minimax estimators, and hence the lack of knowledge about how well minimax estimators behave in particular cases, Hodges and Lehmann (1950, 1951, 1952) embarked on a program to work out minimax estimators in several examples. In particular, these works are concerned with the admissibility of minimax estimators and connections with Bayes estimators. In the first of these papers, Hodges and Lehmann set out to find a minimax estimator for the probability of success p in a binomial experiment. Using the beta conjugate prior for p with parameters α and β, the Bayes estimator turns out to be $(\alpha + X)/(\alpha + \beta + n)$,

J. Rojo
Professor of Statistics, Department of Statistics, Rice University, Houston, TX 77005
e-mail: jrojo@rice.edu

J. Rojo (ed.), *Selected Works of E. L. Lehmann*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1412-4_1, © Springer Science+Business Media, LLC 2012

where X is the number of successes in n Bernoulli trials. The risk function of this estimator is easily computed and, for $\alpha = \beta = \sqrt{n}/2$, has constant risk. Thus, the estimator defined by $(\bar{X}\sqrt{n} + 1/2)/(1 + \sqrt{n})$, where $\bar{X} = X/n$, is minimax. Admissibility of the minimax estimator then follows from the uniqueness of the Bayes estimator. Other strategies for finding minimax estimators in the nonparametric and prediction settings are also discussed. For example, in the nonparametric case, the paper presents an earlier version of the following result: Let $X \sim F$, with $F \in \mathcal{F}$. Suppose that $\delta(X)$ is minimax for $g(F)$ when $F \in \mathcal{F}_{\prime}$ and $\mathcal{F}_{\prime} \subset \mathcal{F}$. If in addition,

$$\sup_{F \in \mathcal{F}_{\prime}} R(F, \delta(X)) = \sup_{F \in \mathcal{F}} R(F, \delta(X)), \tag{1}$$

then $\delta(X)$ is also minimax over $\mathcal{F}$. (See Lehmann and Casella (1998)). Assuming that $X_1, X_2, \ldots, X_n$ are independent with a joint distribution in some family $\mathcal{F}$, and if the family $\mathcal{F}$ contains $\mathcal{F}_{\prime}$, the family of joint distributions of n independent and identically distributed Bernoulli random variables with probability of success $p, 0 < p < 1$, then as discussed above, the minimax estimator for p when restricting attention to $\mathcal{F}_{\prime}$ is $(\bar{X}\sqrt{n} + 1/2)/(1 + \sqrt{n})$. That this estimator remains minimax for the larger class of distributions $\mathcal{F}$, follows after showing that (1) holds. The concept of complete classes having been formalized by Wald in (1950), the paper also shows that, for convex loss functions, the class of nonandomized estimators is essentially complete.

Minimax estimators are not typically Bayes estimators. Rather, they arise as limits of Bayes estimators and, hence, the trick of finding a prior distribution for which the Bayes estimator has constant risk does not usually work. Therefore, in Hodges and Lehmann (1951) a different approach is used to find minimax and admissible estimators. Using the Cramer-Rao lower bound, results are provided to obtain admissible/minimax estimators with respect to the loss function $L(\delta, \theta) = h(\theta)(\delta - \theta)^2$, where $h \geq 0$. The method requires the solution of a differential inequality involving the lower bound for the Mean Squared Error. Extensions were also presented to certain sequential problems.

Motivated by the observation that Bayes estimators are computed assuming complete knowledge of a prior distribution, and minimax estimators correspond to Bayes estimators with respect to a least favorable (or a sequence of least favorable) distribution(s), Hodges and Lehmann (1952) considered, instead, an approach based on *restricted Bayes solutions*. An estimator δ is said to be a *restricted Bayes estimator* with respect to a prior λ if it minimizes $\int R_\delta(\theta) d\lambda(\theta)$ subject to $R_\delta(\theta) \leq C_0$, for all θ, where $C_0 > C^*$, and C^* represents the maximum of the risk of the minimax estimator, and $R_\delta(\theta) = \int R(\delta(X), \theta) dF_\theta(x)$ represents the risk function of δ. It is seen that this is equivalent to choosing a δ^* that minimizes, instead, $\rho \int R_\delta(\theta) d\lambda(\theta) + (1 - \rho) \sup_\theta R_\delta(\theta)$ for some $0 \leq \rho \leq 1$. Conditions are discussed for the existence of restricted Bayes estimators and several examples are provided that illustrate the method.

In Lehmann (1951), in the spirit of Wald's decision theory, Erich proposed a general concept of unbiasedness that is associated with the loss function. Thus, an estimator δ is *L-unbiased*, with respect to the loss function $L(\delta, \theta)$, provided that

$E_\theta(L(\delta, \theta)) \leq E_\theta(L(\delta, \theta^*))$ for all $\theta \neq \theta^*$. It turns out that this reduces, when the loss is squared error, to the usual (mean) unbiasedness concept of David and Neyman (1938). When the loss is absolute error, L-unbiasedness is equivalent to the usual median-unbiasedness concept of Brown (1947). And, for appropriate loss functions, this concept reduces to the usual concepts of unbiased tests and unbiased confidence sets. Rojo (1984) studied the existence of L-unbiased estimators for many other types of loss functions.

Continuing with his pursuit of results dealing with minimaxity and admissibility, Lehmann (1952) addresses the issue of the existence of least favorable distributions. Wald (1950) had provided conditions – albeit too restrictive – for the existence of least favorable distributions that required compactness of the parameter space. In this paper, Erich did away with compactness of the parameter space in the case of hypothesis testing and, more generally, in the case where only a finite number of decisions are available. In Lehmann and Stein (1953), the most powerful invariant test for testing one location parameter family against another is shown to be admissible.

In Bahadur and Lehmann (1955), the relationship between the existence of a necessary and sufficient field and the existence of a necessary and sufficient statistic is studied. In these early papers necessary and sufficient meant minimally sufficient, and a field meant a $\sigma-$field. The concepts of minimal sufficient σ-fields and minimal sufficient statistics first appeared in work of Halmos and Savage (1949) and Bahadur (1954). Detailed methods for constructing minimally sufficient statistics can be found in Lehmann and Scheffé (1950), and in Bahadur (1954). But the question, for example, first raised in Bahadur (1954) of whether the existence of a minimally sufficient $\sigma-$field always implies the existence of a minimally sufficient statistic had not been dealt with. In Bahadur and Lehmann (1955), some light is shed on this problem. The paper provides an example of a minimally sufficient $\sigma-$field that cannot be induced by a statistic. However, the question still remained unanswered. Additional work by Pitcher (1957) provided an example of a family of probability measures for which no minimally sufficient $\sigma-$field and no minimally sufficient statistic exists. The question remained: Does the existence of a minimally sufficient $\sigma-$field always implies the existence of a minimally sufficient statistic or vice versa? The first answer to this question appeared in Landers and Rogge (1974): The existence of a minimally sufficient statistic is neither necessary nor sufficient for the existence of a minimally sufficient $\sigma-$field.

Lehmann (1966) provides an alternative proof to a theorem of Bahadur and Goodman (1952) that shows that the "natural" rule for selecting the best population, out of k populations, is optimal. Other optimality properties of various other procedures proposed in the literature are obtained as a consequence of the new proof. Thus, for example, as a consequence of the results provided in this paper, the optimality of the selection procedures in Bechhofer (1954), Bechhofer *et al.* (1959), and Bechhofer and Milton (1954) are seen to be optimal, when the best population is sufficiently better than the second best, in the sense of maximizing the minimum probability of a correct decision.

References

[1] G. W. Brown. On Small-Sample Estimation. *Ann. Math. Statist.*, Vol. 18, No. 4, pp. 582-585, 1947.

[2] R. R. Bahadur and L. A. Goodman. Impartial Decision Rules and Sufficient Statistics. *Ann. Math. Statist.*, Vol. 23, No. 4, pp. 553-562, 1952.

[3] R. R. Bahadur. Sufficiency and Statistical Decision Functions. *Ann. Math. Statist.*, Vol. 25, No. 3, pp. 423-462, 1954.

[4] R. R. Bahadur and E. L. Lehmann. Two comments on sufficiency and statistical decision functions. *Ann. Math. Statist.*, Vol. 26, pp. 139-142, 1955.

[5] R. E. Bechhofer. A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with known Variances. *Ann. Math. Statist.*, Vol. 25, No. 1, pp. 16-39, 1954.

[6] R. E. Bechhofer and M. Sobel. A Single-Sample Multiple Decision Procedure for Ranking Variances of Normal Populations. *Ann. Math. Statist.*, Vol. 25, No. 2, pp. 273-289, 1954.

[7] R. E. Bechhofer, S. Elmaghraby, and N. Morse. A Single-Sample Multiple-Decision Procedure for Selecting the Multinomial Event Which Has the Highest Probability. *Ann. Math. Statist.*, Vol. 30, No. 1, pp. 102-119, 1959.

[8] F. N. David and J. Neyman. Extension of the Markoff Theorem on Least Squares. *Statistical Research Memoirs*, Vol. II, London, pp. 105-116, 1938.

[9] P. R. Halmos and L. J. Savage. Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics. *Ann. Math. Statist.*, Vol. 20, No. 2, pp. 225-241, 1949.

[10] J. L. Hodges, Jr. and E. L. Lehmann. Some problems in minimax point estimation. *Ann. Math. Statist.*, Vol. 21, pp. 182-197, 1950.

[11] J. L. Hodges, Jr. and E. L. Lehmann. Some applications of the Cramer-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.*, University of California Press, pp. 13-22, 1951.

[12] J. L. Hodges, Jr. and E. L. Lehmann. The use of previous experience in reaching statistical decisions. *Ann. Math. Statist.*, Vol. 23, pp. 396-407, 1952.

[13] D. Landers and L. Rogge. Minimal Sufficient σ-Fields and Minimal Sufficient Statistics. Two Counterexamples. *Ann. Math. Statist.*, Vol. 43, No. 6, pp. 2045-2049, 1972.

[14] E. L. Lehmann. On families of admissible tests. *Ann. Math. Statist.*, Vol. 18, pp. 97-104, 1947.

[15] E. L. Lehmann and H. Scheffé. Completeness, similar regions and unbiased estimation. Part I. *Sankhya*, Vol. 10, pp. 305-340, 1950.

[16] E. L. Lehmann. A general concept of unbiasedness. *Ann. Math. Statist.*, Vol. 22, pp. 587-591, 1951.

[17] E. L. Lehmann. On the existence of least favorable distributions. *Ann. Math. Statist.*, Vol. 23, pp. 408-416, 1952.

[18] E. L. Lehmann and C. Stein. The admissibility of certain invariant statistical tests involving a translation parameter. *Ann. Math. Statist.*, Vol. 24, pp. 473-479, 1953.

[19] E. L. Lehmann. On a theorem of Bahadur and Goodman. *Ann. Math. Statist.*, Vol. 37, pp. 1-6, 1966.

[20] E. L. Lehmann and G. Casella. *Theory of Point Estimation,* 2^{nd} Edition. Springer-Verlag, 1998.

[21] E. L. Lehmann. *Reminiscences of a Statistician: The Company I Kept*, Springer, 2008.

[22] T. S. Pitcher. Sets of Measures not Admitting Necessity and Sufficient Statistics or Subfields. *Ann. Math. Statist.*, Vol. 28, No. 1, pp. 267-268, 1957.

[23] J. Rojo. *On Lehmann's General Concept of Unbiasedness and the Existence of L-unbiased Estimators*. Ph.D. Thesis, Department of Statistics, University of California at Berkeley, 1984.

[24] A. Wald. *Statistical Decision Functions*. John Wiley & Sons, 1950.