

The First Erich L. Lehmann Symposium – Optimality

The First Erich L. Lehmann Symposium was held in Guanajuato, México during May 23 – 25, 2002, and was hosted by the Centro de Investigaciones Matemáticas (CIMAT). The idea to develop the Symposia following the model of the influential, now defunct, Berkeley Symposia, originated while Victor Pérez Abreu and I dined in one of the many beautiful Haciendas in and around Guanajuato.

Our goal was to develop a series of symposia that will attempt to examine the role that Optimality can play, or should play, in modern statistics. Due to the advent of high throughput data collection technology and the parallel development of computing power to analyze such data, it often happens that statistical theory gives way to raw computing power. Although most of the new exciting statistical methodologies have provided tools to make headway in many important scientific problems, a need to generalize and systematize this knowledge is now quite evident.

I approached Erich with the idea during the Joint Statistical meetings in Atlanta. I thought Erich was the right person to help us in bringing the idea to fruition. Not only is Erich a direct professional descendant of Jerzy Neyman, but his lifetime contributions to statistics – research, teaching, and human resources development – provide the ideal stage for the Symposia. [Erich’s lifetime activities have been documented in more detail in an introduction for his Noether award, and these will be the topic of a paper, in progress, that will appear in the Volume of the Second Lehmann Symposium.] Characteristically, Erich refused. The reader should know that one of Erich’s strong virtues is modesty. He would not have any part in this. Fortunately, after several pleas to reconsider, he accepted.

The time available for planning was short. All invited speakers and contributed session speakers were kind to prepare in such a short notice. The First Erich L. Lehmann Symposium was a success. CIMAT was a wonderful host – as they always are. The environment provided by CIMAT and the surrounding colonial structures was stimulating scientifically and culturally. Social activities included a typical “callejoneada” consisting of walking, while drinking wine, through the narrow alleys of Guanajuato surrounded by colonial structures dating back to the 17th century, and led by a “estudiantina”. A tour of

historical San Miguel Allende, Dolores Hidalgo, and Atotonilco, -- all important historical places during the Mexican war for Independence -- was magnificent.

The nine articles presented here are a subset of all the papers presented during the Symposium. All papers have been refereed.

In the first article, *Optimality and Symposia: Some History*, Erich L. Lehmann provides an interesting account of the development of the Neyman-Pearson lemma based on existing letters from Neyman to Pearson. A photocopy of one of Neyman's letters, where he communicates to Pearson his formulation of the simple vs simple hypothesis testing problem as an optimality problem, is included. In addition, Lehmann provides insights on Neyman's life and puts in perspective the history and impact of the Berkeley Symposia.

In the second paper, Juliet P. Shaffer provides a thorough historical survey of the important problem of multiple testing, connecting some modern approaches to earlier work of Lehmann. The paper focuses on cases where only a finite number of parameters are involved.

The third paper by Javier Rojo deals with the nonparametric estimation of distribution functions that are stochastically ordered. As the empirical distribution functions do not satisfy the constraint, new estimators are derived that satisfy the constraint and have better mean squared error properties than the empirical distribution function. An asymptotic test for the equality of two distribution functions against an alternative of stochastic order is also provided, and the asymptotic distribution theory for the estimators is delineated.

Chun-Houh Chen and Ker-Chau Li utilize the Zip code data to illustrate their proposal for a three-way sub-classification algorithm that reduces a k -class classification problem to $k(k-1)/2$ 2-class problems. This approach provides ease of computation and good error rates.

In the paper *A SiZer analysis of IP Flow Start Times*, Stephen Marron *et al* utilize the SiZer technique – Chaudhuri and Marron (1999) -- to assess the fit of various models to the point process of internet traffic flow start times. It is concluded that a clustered Poisson process provides a better fit than a homogeneous Poisson process and a Weibull waiting process.

Nicolas Hengartner and Marten Wegkamp's paper establishes a stochastic approximation for studentized M-estimators in terms of a certain U-statistic. This result is obtained under weaker technical conditions than those available in the literature. Under somewhat stronger conditions, expansions are developed that connect the bootstrap approximation to the distribution of the studentized M-estimator.

Dennis Cox considers the optimality of unbiased predictors for stochastic processes. When the process is Gaussian, the best linear predictor remains optimal in the class of unbiased predictors; if the process is log-Gaussian, the standard predictor is optimal in the class of all unbiased predictors when using a weighted mean squared error criterion.

David Sprott argues in his paper, *What is Optimality in Statistical Inference*, that extraneous optimality criteria should not supersede inference based only on the likelihood function. A case is presented for the relative likelihood function as a main tool for statistical inference and several examples are developed to illustrate this approach

The last article by Díaz-Francés and Sprott considers the problem of estimating two proportions under the constraint that one is larger than the other. The approach is based on examining the observed likelihood function coming from both samples. This is in sharp contrast with existing proposals that recommend disregarding the information arising from one of the samples.

Armando Domínguez provided substantial technical support in bringing the document to its final form. Armando's efforts are gratefully acknowledged.

Finally, we would like to thank Rice University and CIMAT for allowing us time to organize the symposium and edit this volume. The local organizing committee in Guanajuato did an excellent job. CIMAT, under the direction of Victor Pérez Abreu, provided financial support for the First Symposium.

Javier Rojo
Rice University
Houston, TX