HAPPY BIRTHDAY CHARLES

MY TALK IS TITLED:

Charles Stein's Research Involving Fixed Sample Optimality,
Apart from Multivariate Normal Minimax Shrinkage
[*aka*: Everything Else that Charles Wrote]

Lawrence D. Brown
Statistics Department, University of Pennsylvania

March 22, 2010

Building and Understanding Frequentist Optimality

1. A two-sample test for a linear hypothesis whose power is independent of the variance. (1945).

0. (with Hunt, G. A.) Most stringent tests of statistical hypotheses (1946??), unpublished

2. A note on cumulative sums. (1946).

3. (with Wald, A.) Sequential confidence intervals for the mean of a normal distribution with known variance. (1947).

4. (with Lehmann, E. L.) Most powerful tests of composite hypotheses. I. Normal distributions. (1948)

5. Some problems in sequential estimation. (1949).

6. (with Lehmann, E. L.) On the theory of some nonparametric hypotheses. (1949).

7. Unbiased estimates with minimum variance. (1950).

8. (with Lehmann, E. L.) Completeness in the sequential case. (1950).

9. A property of some tests of composite hypotheses. (1951).

10. (with Lehmann, E. L.) The admissibility of certain invariant statistical tests involving a translation parameter. (1953).

10a. On tests of certain hypotheses invariant under the full linear group (abstract). (1955).

11. A necessary and sufficient condition for admissibility. (1955).

12. The admissibility of Hotelling's T^2-test. (1956).

For Example

4. (with Lehmann, E. L.) Most powerful tests of composite hypotheses. I. Normal distributions. (1948)

- Contains
    1. Useful <u>methods</u> of investigating properties of standard tests
    2. Some interesting and <u>encouraging</u> results
    3. <u>Some</u> slightly <u>unintuitive</u> (and possibly discouraging) <u>ones</u>

Also to be noted

7. Unbiased estimates with minimum variance. (1950).
8. (with Lehmann, E. L.) Completeness in the sequential case. (1950).
- Involve attempts to unify the theory by incorporating together both sequential and non-sequential formulations. This unifying perspective also appears elsewhere in Stein's work and influenced others, including Kiefer (1957) and me (1966?) and undoubtedly influenced LeCam (1953? +). (Dave Siegmund will discuss Charles' contributions to sequential analysis, *per se*.)

A Multivariate Result using a
Valuable Exponential Family Technique

12. The admissibility of Hotelling's T^2-test. (1956)

- In (1955) Birnbaum had proved:

  Let $X \sim N_n(\theta, I)$. Consider testing $H_0 : \theta = \mathbf{0}$ vs $H_0 : \theta \neq \mathbf{0}$. Then a test is admissible if and only if its acceptance region is a convex set (up to sets of measure 0).

  Stein extended this idea to prove Hotelling's T^2-test is admissible.

- This was a foray into the realm of truly multivariate distributions. (Joe Eaton will detail some of Charles' accomplishments in this realm.) AND
- Charles generalized Birnbaum's proof into a general method that was built on ….

One of several

    Stein's Lemma(s): X comes from an n-dimensional natural exponential family with natural parameter vector θ. The setting in 2-d is
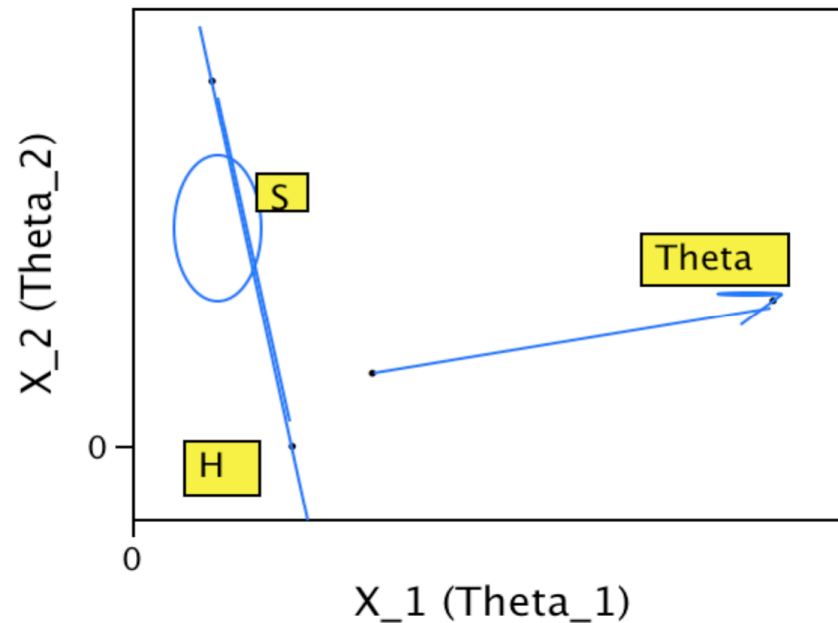
H is a half-space and S is a possibly small set that pokes into the outside of H. $\theta \to "\infty"$ along the outward normal from H. Then

$$\lim_{\theta \to "\infty"} \left\{ P_\theta(S) \big/ P_\theta(H) \right\} = \infty.$$

From Invariance to Admissibility

Papers:

0. (with Hunt, G. A.) Most stringent tests of statistical hypotheses (1946??),

10. (with Lehmann, E. L.) The admissibility of certain invariant statistical tests involving a translation parameter. (1953).

11. A necessary and sufficient condition for admissibility. (1955).

&

13. Efficient nonparametric testing and estimation. Third Berk Symp. (1956).

14. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Third Berk Symp. (1956). [Sections 3 & 4]

15. The admissibility of Pitman's estimator of a single location parameter. (1959).

16. An example of wide discrepancy between fiducial and confidence intervals. (1959).

17. Multiple regression. Stanford University Press. (1960).

18. (with James, W.) Estimation with quadratic loss. 4th Berkeley Symp (1961). [Sections 4 and 5]

&

23. Approximation of improper prior measures by prior probability measures. (1963 – or 1964?)

From Invariance to Admissibility

- In (1939) Pitman laid out a schema for finding best invariant tests and estimators in location and scale problems:

For example, if $X_i \sim_{iid} f(x - \theta)$ and an estimate δ of θ is wanted under an invariant loss, $L(\delta - \theta)$, then invariant estimators have the form $\delta = \hat{X} + c(Y)$ where $Y = X_2 - X_1, ..., X_n - X_1$ and $\hat{X}$ denotes some convenient equi-variant functional such as $\text{med}\{X_i\}$ or $\bar{X}$.

- The form is simple and seductive. The Hunt-Stein theorem (1946) shows that best invariant procedures for this setting are minimax.

{Joe Eaton will discuss this paper and the sequel minimax developments. My suspicion is that the original H-S only proved minimaxity for the testing setup, but it's applicability also to estimation was realized only a few years later.}

<center>Admissibility in One Dimension</center>

1. Wald (1939) claimed admissibility for such best invariant estimators, but his proof contains a "lacuna" (quote is from Peisakoff (1950))
2. Three different proofs of admissibility were published in 1951. Two [A & B, below] were only for the normal mean problem (known variance); the third was for an artificial looking discrete problem:

> A. Blyth – using a sequence of proper priors.
> B. Hodges and Lehmann – using the Cramer-Rao inequality
> C. Blackwell – using truncations of the uniform prior

- Something {Logical parsimony (?), Statistical buoyancy (?), &/or Mathematical ignorance(?)} suggested admissibility should hold in general.

- Lehmann and Stein (1953) proved admissibility in a fairly general **1-d** testing setting. (Testing between two different **1-d** location families.)
Used Blackwell's method! Need a *seemingly* unnecessary moment condition; but see Fox& Perng (1969).

- Stein (1955) examined Blyth's method. He showed (under mild conditions) that **if** admissibility holds, **then** Blyth's method could – in principle – be used to prove it. [See also later work by Farrell, by Brown, etc]
BUT Blyth's method requires invention of a "good" sequence of proper priors. So, failure to succeed with the method doesn't prove inadmissibility!

Admissibility in 2-d

- Proof using C-R Inequality appears in 3rd Berk. Symp. - along with proof of inadmissibility in 3+ dimensions.

- **Which came first?** [I don't know, but Charles does.]

- Proof makes pretty clear that admissibility should fail for any dimension > 2
    - AND IT DOES FAIL as the first part of the paper shows

{Carl Morris will talk about this. And so will Brad Efron.}

SO

More about ADmissibility

- Stein (1959) applies Blyth's method to prove a general admissibility result for best invariant estimators in the 1-d location problems.
- Sections 4 and 5 of the 4th Berk. Symp. use Blyth's method to prove admissibility for the general 2-d location problem with squared error loss.
  - A generalization for other losses is in Brown and Fox (1968)
  - Stein's proof requires a clever choice for the sequence of parameters. At the time, to all but Charles this seemed inspired from !!!! . There is an attempt to explain this sequence in Stein's tech report (1964?). And (I hope) this is further clarified in Brown (1971). Stein's prior sequence remains useful, as in Brown and Han (2010).

- These general admissibility results assume finiteness of some extra moments that seem extraneous to the problem. But these were shown to be needed in Brown (1966) by an artificial example and by a natural one in Fox (1981). Is this somehow related to the geometrical explanation for inadmissibility?

Additional (Miscellaneous) Comments

1.   Regarding

26. (with Portnoy, S.) Inadmissibility of the best invariant test in three or more dimensions. (1971).

This is about inadmissibility *(so probably isn't in my domain for this discussion)*.

It closes a gap in the conceptual circle by showing that inadmissibility in 3+ dimensions also holds for location invariant testing problems.

2.   Regarding

17. Multiple regression. Stanford University Press. (1960).

This clarifies the intimate connection between the location parameter problem and the more prominent multiple regression setting. Most of the specifics are about inadmissibility *(and so not in my domain)* but the generalities and insights are relevant to all statistical applications of regression.

3. Regarding

21. (with Giri, N. and Kiefer, J.) Minimax character of Hotelling's T^2 test in the simplest case. (1963).

This is a multivariate invariance issue (*and so belongs to Joe*). It gives an extremely partial answer to an issue Charles had raised earlier. (*So far as I know no one has yet answered the general question posed there.*) Since I was a grad student at the time Jack Kiefer was working with Charles on this, I know how difficult and resistant to solution the problem is, and how proud they were to have gotten this far, and how frustrated not to be able to progress further.

4. Regarding

22. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. (1964).

This is also about inadmissibility, but involves the estimation of only one parameter (the variance of a normal sample when the mean is unknown). (*So maybe it belongs to me!*) The loss function here is a form of Kullback-Liebler loss that others (including me in Brown (1967)) refer to as "Stein's loss". The loss appears first in Section 5 of the 4th Symp. paper in the multivariate case. (*So maybe it belongs in Joe's discussion?*)

5.   Geometry *vs* empirical Bayes:

The 3rd Symp. paper starts with a beautiful, totally convincing heuristic argument that the usual estimator must be inadmissible when the dimension is large. We (now) recognize this as being about the geometry of higher dimensional space. Hence it's a "mathematical" argument.

In

19. Confidence sets for the mean of a multivariate normal distribution. (1962).

In two sentences Charles presents a new perspective on inadmissibility that is statistical rather than geometric. This is what we now recognize as the

parametric empirical Bayes approach

(*So the rest of this story belongs to Carl and Brad*)
Charles writes that this shows the inadmissibility is "not as substantial a break with current practice as some people, including the author, have previously assumed."

About Insights, Intuition and Idea-Sharing

- Charles mind has always been loaded with insights and ideas. He has been generous about sharing these in person and at each opportunity in his papers:
- Many of these have been pursued and utilized by others. Some raise still vibrant questions that should yet be pursued further.
- Some examples are the geometry, empirical Bayes and multiple regression discussions I've mentioned. To save time and space let me emphasize this quality within Stein's publications by pointing to a remarkable collection of comments in his 4th Berk Symp masterpiece. Charles describes this as

   "In section 4 several conjectures and unsolved problems … have been mentioned" [*and gems of insight have been buried in the descriptions*]. "Some other problems are listed below."