Discussion of Ennis (2008):

DTU Informatics
Statistics

Hypothesis Testing for Equivalence defined on Symmetric Open Intervals

Per Bruun Brockhoff
& Rune H.B. Christensen
DTU Informatics,
Statistics.

Main issue:

Hoenig and Heisey (2001), *The American Statistician:* The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis.

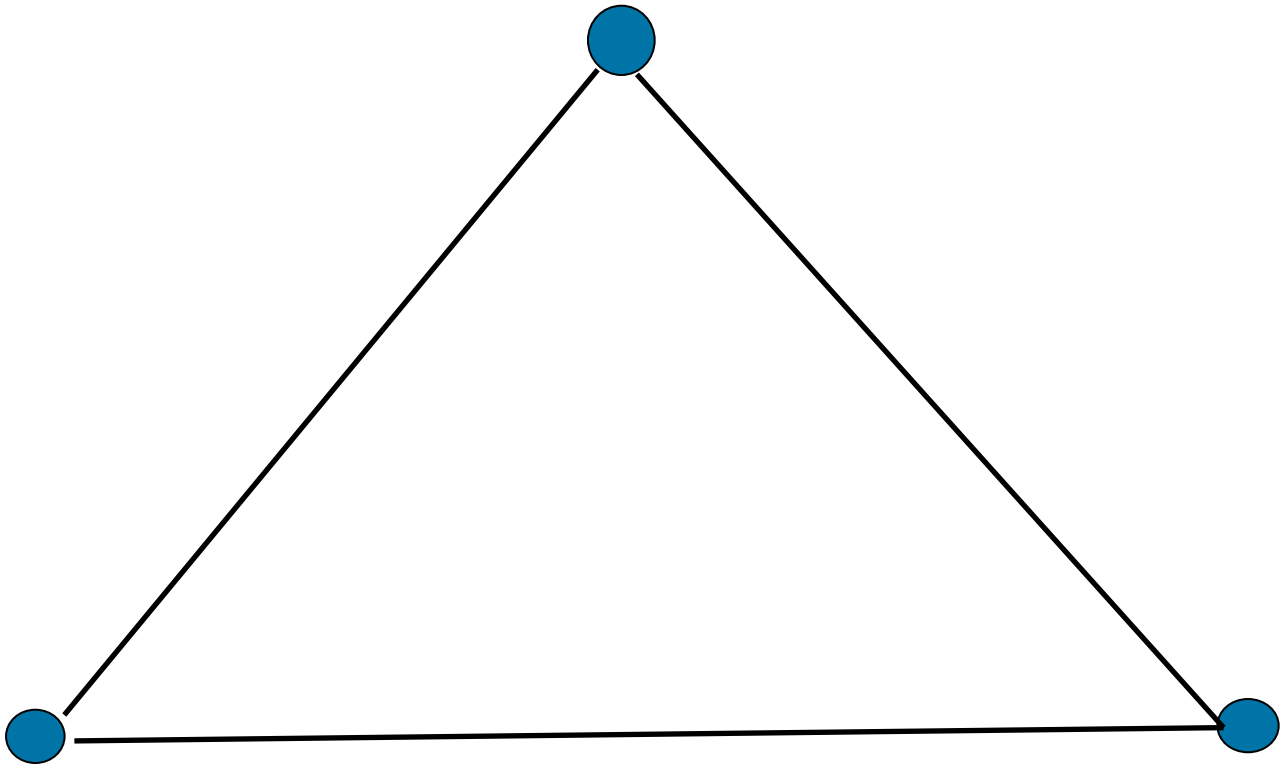
DO NOT use power for data analysis!!
(Instead: Use Confidence regions!)

ONLY use power in planning phase!!

(By the way: support simple TOST for equivalence testing)

Bradley Efron (1998):

(Statistical Science, 95-122)



Pawitan (2001):

(In All Likelihood, Oxford University Press)

"Statisticians are probably unique among Scientists with constant ponderings of the foundation of their subject……"

Fisher-Neyman Debate

Neyman: **Stringent approach:** **Uniformly most powerful test**
Unbiasedness
alpha- and d-Admissibility

Berger and Hsu (1996): *"We believe that notions like size, power and unbiasedness are more fundamental than "intuition""*

Fisher: "Common Sense" **No single and clear optimal testing theory! (if NOT Bayesian)**

Use likelihood

Pearlman and Wu (1999)

Statistical Science, 355-381

The Emperor's New Tests: (final paragraph, p. 381)

"In conclusion we have no Grand Unifying Testing Theory to present, no universal mathematical criterion for judging all tests in all testing problems as Berger asks of us, and think it unlikely that such criterion exists. Depending on one's point of view, this nonexistence may be seen as either a strength or a weakness of statistics, but in any event it is better to have no universal criterion than cling to an inappropriate one. We hope that we have alerted statisticians to the dangers inherent in uncritical application of the NP criterion, and , more generally, convinced them to join Fisher, Cox and many others in carefully weighing the scientific relevance and logical consistency of any mathematical criterion proposed for statistical theory."

Equivalence in Sensometrics

• MacRae (1995). Confidence intervals for the triangle test can give assurance that products are similar. *FQP 6, 61-67.*

• Carr (1995). Confidence intervals in the analysis of sensory discrimination tests – The integration of similarity and difference testing. *4th Agrostat, 22-31.*

• Bi (2005). Similarity testing in sensory and consumer research, *FQP 139-149.*

• Bi (2006). Similarity testing using paired comparison method, *FQP 500-507.*

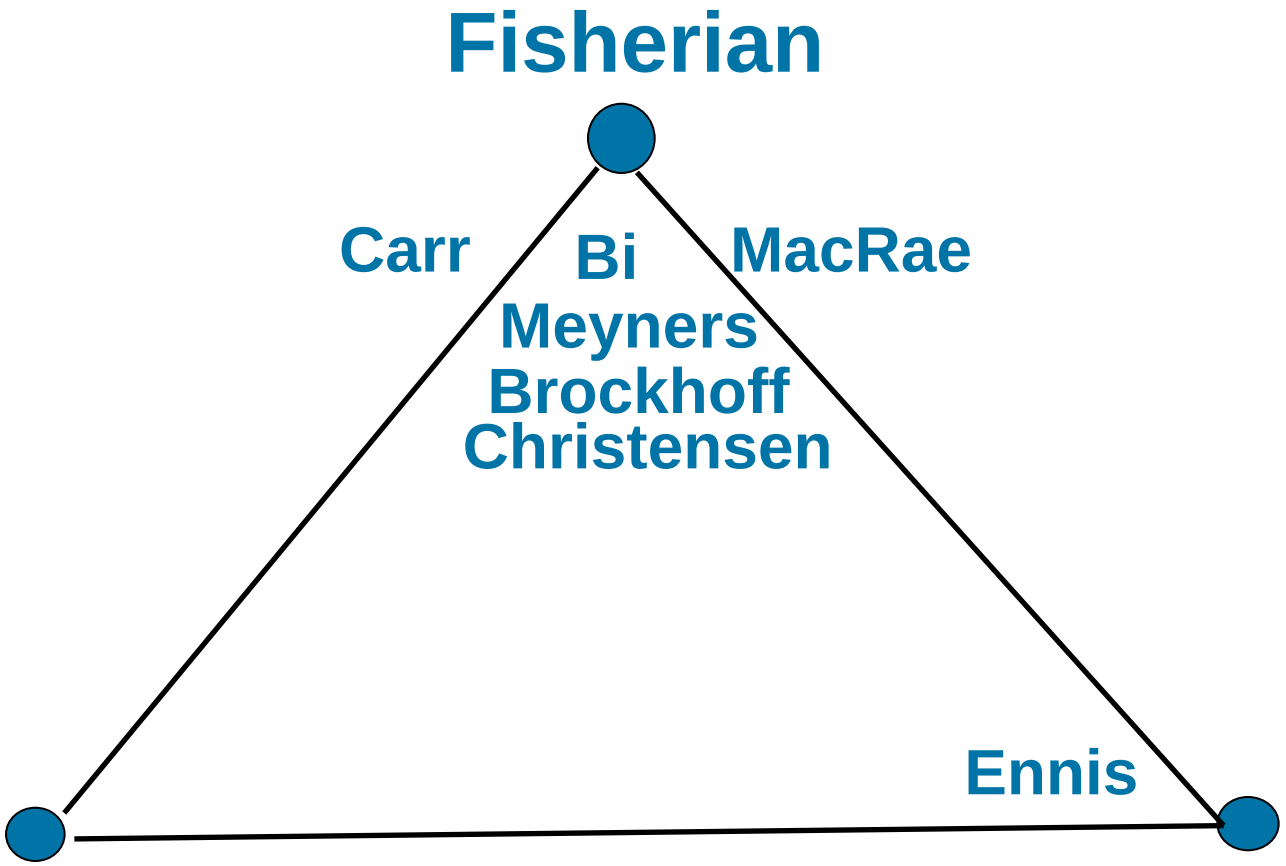
• Meyners (2007). Least equivalent allowable differences in equivalence testing. *FQP 541-547.*

• Bi (2007). Replicated similarity testing using paired comparison method, *J. Sens. Stud. 176-186.*

• Ennis, Bi and Meyners (2008): Response and discussion of Bi (2007), *FQP 344-348.*

• Ennis (2008). Tables for parity testing. *J. Sens. Stud. 80-91.*

Equivalence in Sensometrics



Russel A. Boyles (2008):

(The American Statistician, p. 22-26)

The Role of Likelihood in Interval Estimation

Focus on "*law of likelihood*":

In relation to a model, data supports p* more strongly than p if and only if the likelihood of p* is larger than that of p.

Recommendation: **Use likelihood based Confidence Intervals**

Results for binomial data:

Highest-Likehood intervals perform better than standard AND also better than exact intervals!! (Clopper-Pearson, 1934)

Likelihood based Intervals for d-primes

Brockhoff, P. B. and R. H. B. Christensen (2008). Thurstonian model as Generalized linear models. *Draft for FQP.*

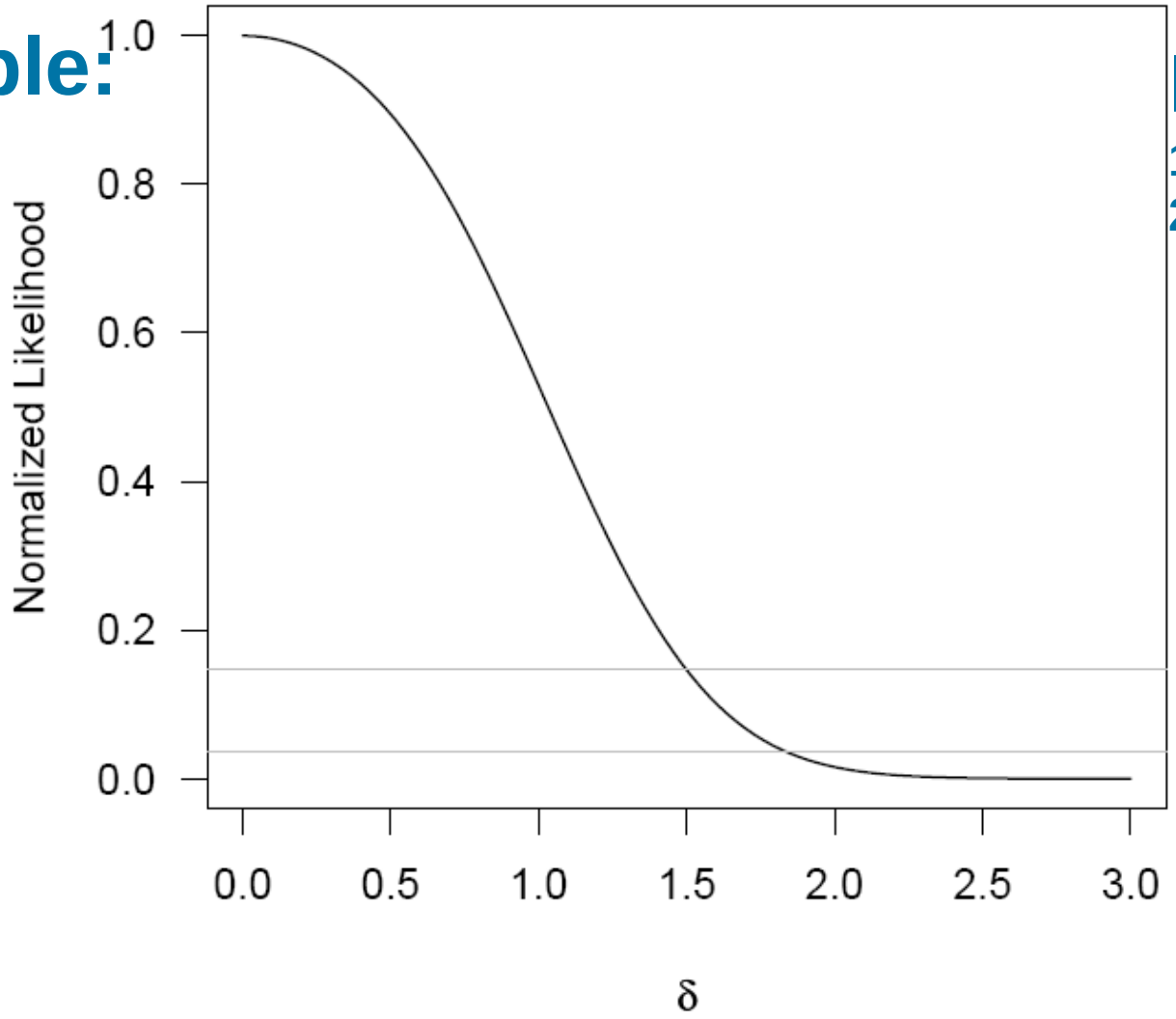
Christensen & Brockhoff (2008c). Estimation and Inference in the Same Different Test, *Draft for FQP*.

Christensen, R. H. B. and P. B. Brockhoff (2008a). The application of Thurstonian models for replicated difference tests. *Draft for FQP*, presented at the Sensometrics 2008 Conference July 20-23, Canada.

Christensen, R. H. B. and P. B. Brockhoff (2008b). SensR: An R-package for Thurstonian modelling of discrete sensory data. R-package, presented at the Sensometrics 2008 Conference July 20-23, Canada.

Likelihood based Intervals:

Duo-Trio example:

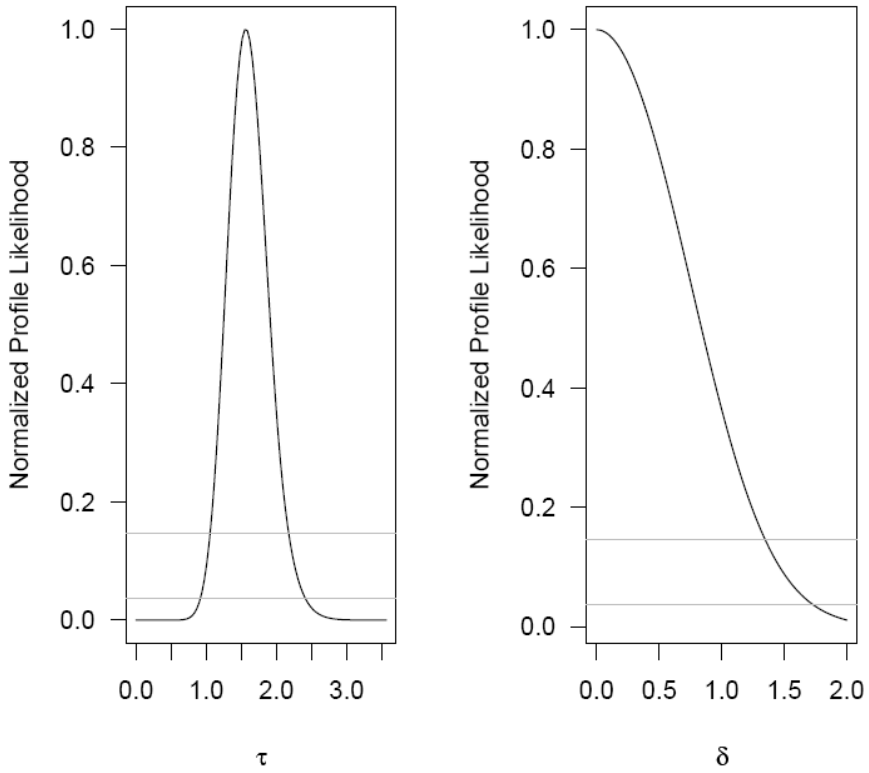


Results:

10 correct out 22 test.

Likelihood based Intervals:

SAME-DIFF example:



Results:

8 same ans. out of 11 same samples

9 same ans. out of 11 diff samples

Figure 3: Profile likelihood of δ and τ for Mark's experiement. Horizontal gray lines denotes 95% and 99% confidence limits.

In conclusion:

Instead of

discussing **COMPLICATED** Neyman-Fisher controversies
for **SIMPLE** sensory equivalence problems

we should use the time and effort

Discussing the **COMPLICATED (and SIMPLE)** models/settings
for the (more) **COMPLICATED** sensory equivalence problems.

(And adopt the Fisherian pragmatism in lettting the confidence region approach do the job for us – integrating nicely models with estimation and difference and similarity testing)