Comment on DM Ennis: Hypothesis testing for equivalence defined on symmetric open intervals

Michael Meyners

Proposed method

- Power increase
- Easily performed
- Applicable also in some non-normal situations

Proposed method – application to

- Complex models?
  (cross-over, mixed models, …)
- Non-parametrics?
- Randomization tests?

- Validity of the chi-squared approximation?
- Choice of c

Relation to "emperor's new tests"

- Rejection region qualitatively the same
- Known relationship between χ^2, t and F distribution
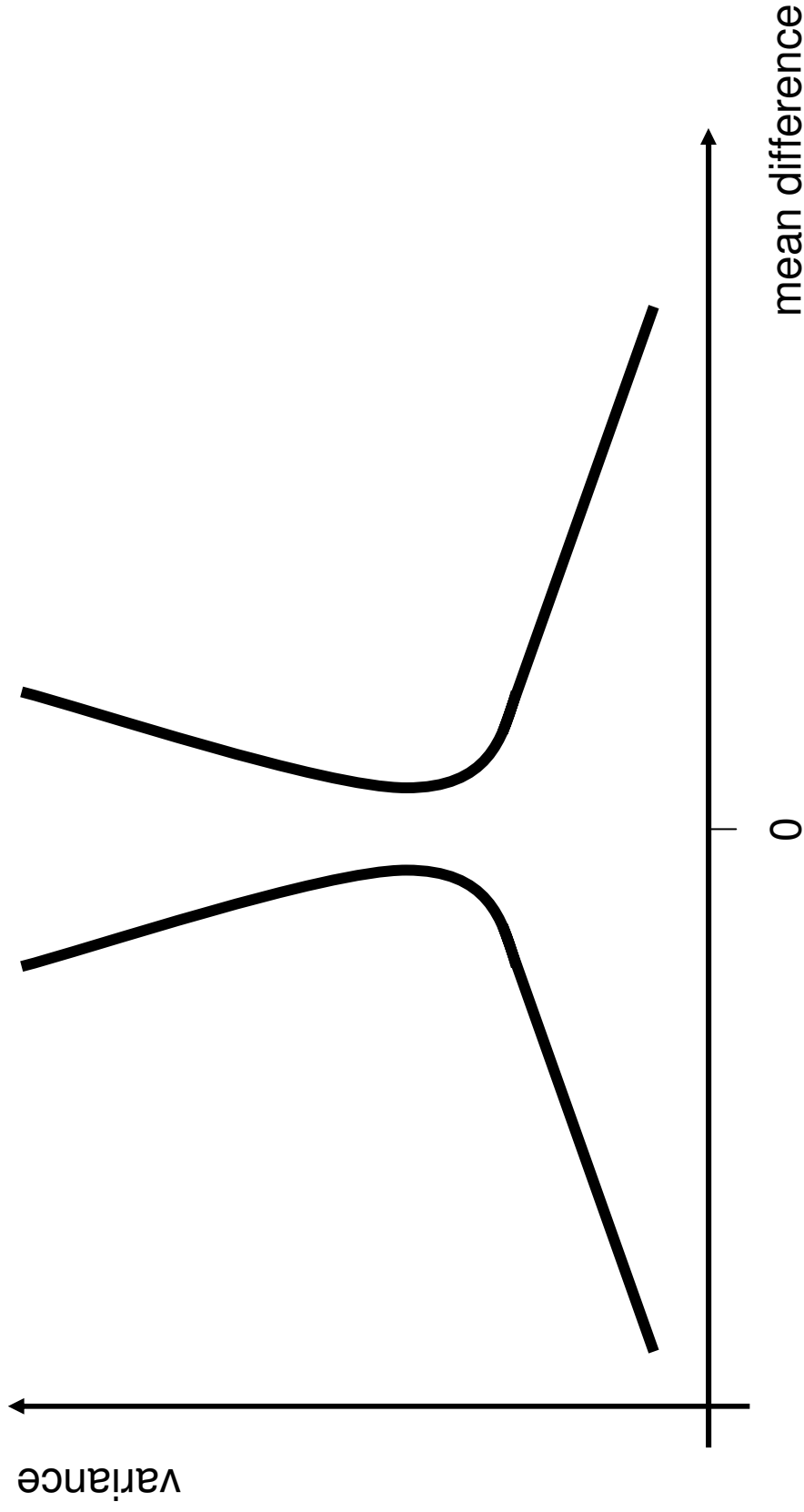
Relation to confidence intervals

- Level α or level 2α?

Comparison to competitors based on criteria alternative to power

- Does not minimize risk under null hypothesis
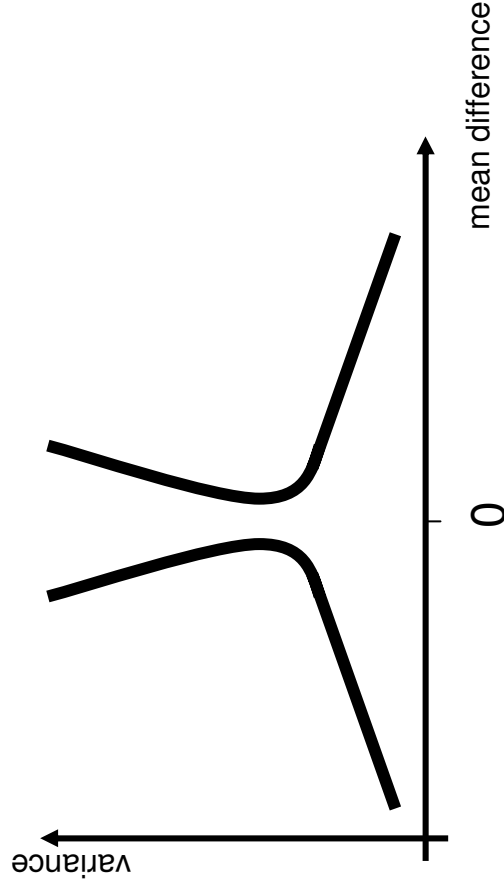
- Shares this property with "emperor's new tests"

Unbounded rejection region (sketch)

variance

mean difference

0

- significant equivalence if test statistic falls into this region

Unbounded rejection region – possible implications

- Reject $|\delta| > 1$ but the estimate for δ is 1.5

- Reject $|\delta| > 1$ but fail to reject either $\delta > 1$ or $\delta < -1$

Definition of equivalence limits

- Increase in power much easier

- Limits are often chosen to narrow (at least in the beginning)

- Impact on power much higher than choice of method

Personal point of view

- Always look at the two one-sided tests (individually)

But

- What is the primary goal ?

Parity, superiority claims

- Power is the most important criterion (possibly even the only one)

Product development

- Confidence in similarity required

- Power relevant for sample size determination

- Type I error rate much more important

Often termed "intuition"

- Rather a different concept

- More relevant in product development