Adaptive Design: Where Are We Now?

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

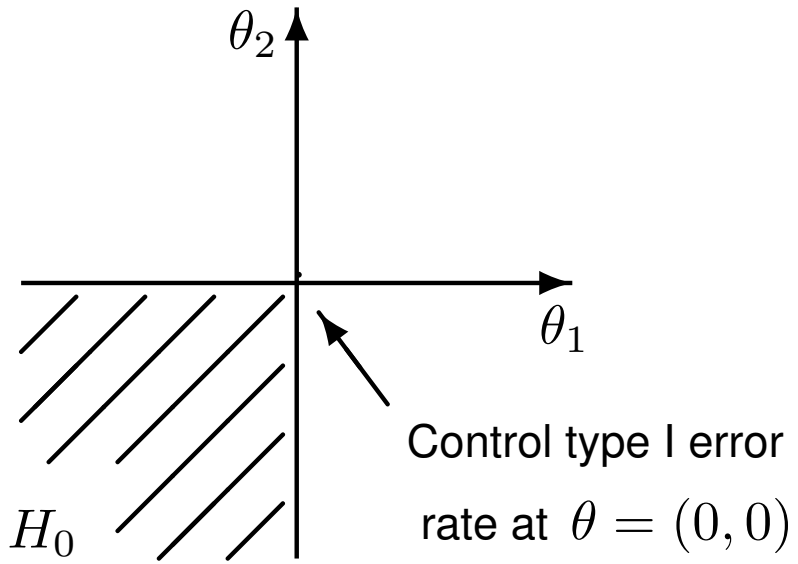
JSM, Washington

4 August 2009

Statistical validity of an adaptive trial

Q: Can the α level of an adaptive design be determined based on simulations?

A: This may not be easy for a complex adaptive procedure with a high dimensional parameter space.



Is the type I error rate controlled over all of H_0?

Counter examples

Controlling type I error at an "intuitively appropriate" point does *not* necessarily control type I error over all of H_0.

Turnbull, Kaspi & Smith (*JSCS*, 1978) define an adaptive selection procedure whose maximum type I error probability is not at the "least favourable configuration" of a non-adaptive design.

Stallard & Friede (*SiM*, 2008) present a group sequential design with adaptive treatment selection. The type I error rate holds if numbers of treatments retained at each stage are pre-fixed but not if they are data-dependent.

In a seamless Phase II/III design, Jennison & Hampson (2009) find an optimal rule for combining data from the two phases with type I error rate α at $\theta = (0, \ldots, 0)$ — but the error rate is *higher* when some θ_i are negative.

Adaptive can be counter-intuitive: Jennison & Turnbull (*SiM*, 2003) show an adaptive design with a non-monotone power curve.

Solutions

We wish to show type I error rate is controlled over a high dimensional H_0 — and an exhaustive check is not feasible.

(1) Theory can sometimes give a complete answer.
Often, use of inequalities leads to conservative procedures.

(2) Use theory to reduce the problem to checking (e.g., by simulation) over a simpler space than the original H_0.

(3) Modify the adaptive procedure so its analysis is more tractable.

Example of (3): Updating sample size in a group sequential test in response to estimates of the response variance:

An error spending design with "information monitoring" (Mehta & Tsiatis, *Drug Inf. J.*, 2001) can give a slightly inflated type I error rate.

In the framework of a group sequential combination test (Lehmacher & Wassmer, *Biometrics*, 1999), type I error rate is guaranteed exactly.

Estimates and confidence intervals after an adaptive trial

Q: What are the requirements for point estimates and confidence intervals?

A: If these are reported as frequentist data summaries, they should possess the appropriate frequentist properties.

Chapter 8 of "Group Sequential Methods with Applications to Clinical Trials" (Jennison & Turnbull, 2000) describes methods for inference after a group sequential test — p-values, confidence intervals, point estimates.

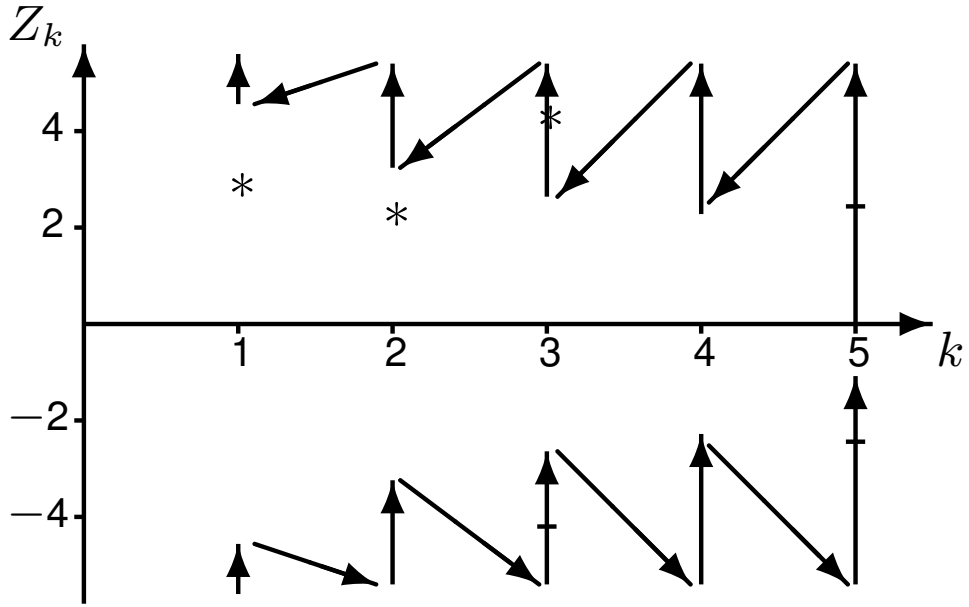
For generalisations to adaptive designs see, e.g.,

Brannath, Posch & Bauer(*JASA*, 2002),

Posch et al. (*SiM*, 2005),

Brannath, Mehta & Posch, (*SiM*, 2009).

Estimates and confidence intervals



Inference on termination of a group sequential test is usually based on an ordering of the sample space.

The “stagewise ordering” is a convenient choice.

Other choices of ordering are possible, so the analysis on termination may not appear as “definitive” as one might like.

Estimates and confidence intervals

Recent work by Brannath and Posch shows joint inferences about multiple treatment effects can require sophisticated constructions in order to deliver useful looking results.

It seems the frequentist framework is struggling to cope with these complexities!

It may be asking too much to try and find an all-purpose summary of multiple parameters.

Some specific research questions will be of prime importance — and these could be tackled directly, e.g., by step-down and gate-keeping procedures for testing multiple hypotheses.