Group Sequential Tests with Data-Dependent Treatment Allocation

Chris Jennison,

Department of Mathematical Sciences, University of Bath, UK

and

Bruce Turnbull,

Department of Statistical Science, Cornell University, Ithaca, NY

MCP 2002,

Bethesda,

5–7 August 2002

http://www.bath.ac.uk/∼mascj

for slides and full references

The Selection Problem

For each "population" or "treatment" $i = 1, \ldots, k$,

$$X_{i1},\, X_{i2},\, \ldots \;\sim\; N(\theta_i,\, \sigma^2), \quad \text{i.i.d.}$$

Aim: To select the population i with the largest mean θ_i.

Method to include:

- Group sequential comparisons
  - early elimination of weak treatments.
- Response-dependent treatment allocation
  - fewer observations on inferior treatments,
  - lower total sample size.

Earlier work

Paulson (*Ann. Math. Statist.*, 1964)

Elimination procedures based on *continuous* sequential comparisons of 2 populations at a time.

Robbins and Siegmund (*JASA*, 1974)

Adaptive sampling for a 2 population comparison with continuous monitoring.

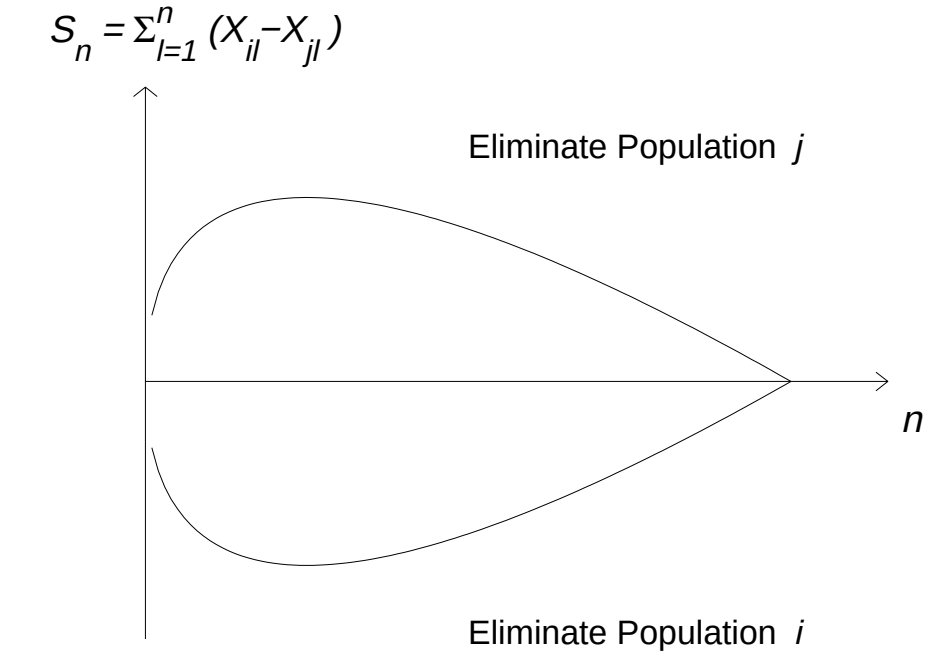
Jennison, Johnstone and Turnbull (*Purdue Symposium*, 1982)

Combining the above.

Update: To take advantage of group sequential tests, error spending, modern computation.

Paulson's procedure

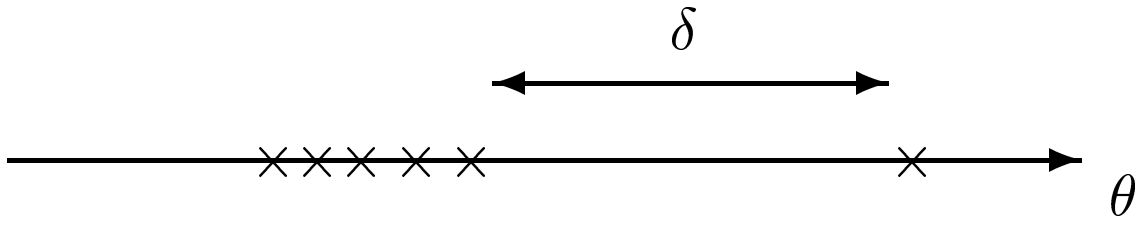
Compare all pairs Treatment i *vs* Treatment j.



If $\theta_i = \theta_j - \delta$, then $Pr\{\text{Pop. } i \text{ eliminates Pop. } j\} = \alpha/(k-1)$.

Paulson's procedure: Probability of Correct Selection

Indifference Zone formulation



Suppose $\theta_i \leq \theta_k - \delta$ for $i = 1, \ldots, k-1$.

Then

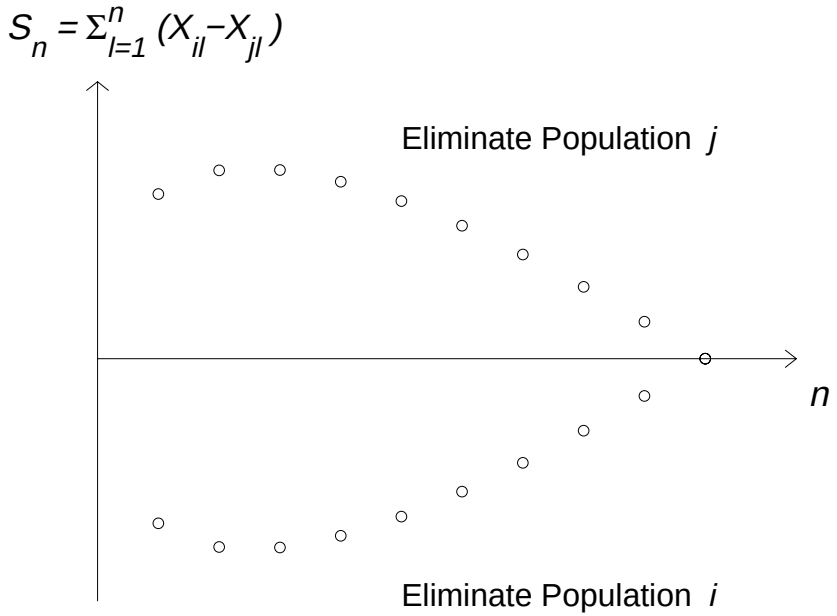
$$Pr\{\text{Pop. } k \text{ is eliminated at some stage}\}$$

$$\leq \sum_{i=1}^{k-1} Pr\{\text{Pop. } i \text{ eliminates Pop. } k \text{ at some stage}\}$$

$$\leq (k-1)\frac{\alpha}{k-1} = \alpha.$$

Paulson's procedure: Group Sequential monitoring

Compare treatments at regular interim analyses.



Choose a group sequential boundary with error rate $\alpha/(k-1)$ at $\theta_i - \theta_j = \pm\delta$ and good early stopping under likely $(\theta_1, \ldots, \theta_k)$.

Adaptive Sampling in Paulson's procedure

Motivation

Observations on the leading population are used in $k-1$ comparisons.

Allocating more observations to the leader can

- Reduce total sample size
- Reduce observations on inferior treatments
  - ethical for medical studies
  - we learn more about better treatments.

Need:

Theory to support adaptive sampling in each pair-wise comparison.

Adaptive Sampling in a Group Sequential Test

Jennison and Turnbull (*Sequential Analysis*, 2001)

For a 2-treatment comparison with

$$X_{1i} \sim N(\theta_1,\, \sigma^2) \quad i = 1, 2, \ldots ,$$

$$X_{2i} \sim N(\theta_2,\, \sigma^2) \quad i = 1, 2, \ldots .$$

At analysis m out of M, with n_{1m} observations on population 1 and n_{2m} on population 2,

$$\widehat{\theta}_1(m) - \widehat{\theta}_2(m) = \bar{X}_1(m) - \bar{X}_2(m) \;\sim\; N(\theta_1 - \theta_2,\, \sigma^2(\tfrac{1}{n_{1m}} + \tfrac{1}{n_{2m}}))$$

$$\sim\; N(\theta_1 - \theta_2,\, \mathcal{I}^{-1}(m)), \quad \text{say.}$$

Adaptive Sampling continued

The score statistic

$$S(m) \;=\; \mathcal{I}(m)(\widehat{\theta}_1(m) - \widehat{\theta}_2(m)) \;\sim\; N((\theta_1 - \theta_2)\,\mathcal{I}(m), \mathcal{I}(m)).$$

Without adaptive sampling, $\{S(1), S(2), \ldots\}$ is distributed as a Brownian motion with drift $\theta_1 - \theta_2$ observed at $\mathcal{I}(1), \mathcal{I}(2), \ldots$.

This remains true if group sizes $n_{1m} - n_{1,m-1}$ and $n_{2m} - n_{2,m-1}$ depend on $\widehat{\theta}_1(m-1) - \widehat{\theta}_2(m-1)$ — but sampling *cannot* depend more generally on $(\widehat{\theta}_1(m-1), \widehat{\theta}_2(m-1))$.

Theory generalises to normal linear models containing θ_1 and θ_2.

This extends Robbins and Siegmund (1974) to the group sequential case.

Adaptive Sampling: Problem 1

Problem 1

With $k \geq 3$, interesting sampling rules do not satisfy

"m th group sizes for populations 1 and 2 depend only on $\widehat{\theta}_1(m-1) - \widehat{\theta}_2(m-1)$".

Solution

- Fix sampling ratios at the start of each group,
- estimate $\theta_i - \theta_j$ within each group of data,
- combine estimates with weights $\propto$ variance^{-1}.

This equates to fitting a linear model with additive "stage" effects — recommended in medical studies to avoid bias from time trends.

Adaptive Sampling: Problem 1

JT (2001) assess performance of 2-treatment tests:

With stage effects in the model, one cannot compensate later on for sub-optimal sampling ratios in early stages. Savings in Inferior Treatment Numbers are reduced by about a half.

- Fitting stage effects to avoid bias from a time trend is reasonable.
- If modelling such a trend is not really necessary, data are being used inefficiently
  - ethically dubious for medical studies
  - at best pragmatic in other applications.

Adaptive Sampling: Problem 2

Problem 2

Information levels for comparing populations i and j

$$\mathcal{I}_{ij}(1), \mathcal{I}_{ij}(2), \mathcal{I}_{ij}(3), \ldots ,$$

depend on the sampling rule, which involves $S_{ij}(1), S_{ij}(2), \ldots .$

Standard group sequential designs, including error spending tests, do not allow such a dependence.

Solution A

Reported studies of such "data-dependent analysis times" show only minor effects on error probabilities — trust these studies and ignore the problem!

Adaptive Sampling: Problem 2

Solution B

Recent designs which "adapt" to observed data offer a precise solution:

Denne (*Statistics in Medicine*, 2001),

Müller and Schäfer (*Biometrics*, 2001).

Procedure

- Set up an error spending test for anticipated $\{\mathcal{I}_1, \mathcal{I}_2, \ldots\}$
- Recursively for $m = 1, 2, \ldots,$
  - At analysis m, compute conditional error probabilities given $S(m)$
  - Run stages $m+1$ to M as an error spending test with this conditional error.

A sampling rule (JJT, 1982)

In comparing $N-1$ populations with a control, the most efficient allocation is

$\sqrt{N-1}$ observations on the control to

1 observation on each other population.

Adaptive rule:

At stage m, with N_m non-eliminated populations, sample

$\sqrt{N_m-1}$ observations on the leading population to

1 observation on each other population.

An updated procedure

Eliminate populations using Paulson's pair-wise comparisons.

Run these comparisons as error spending group sequential tests.

a) Base tests on overall population means (cf JJT, 1982)

Sample in stage m to achieve ratios $\sqrt{N_m} : 1 : \ldots : 1$ of total observations on the N_m surviving populations.

b) Combine stage-wise estimates of each $\theta_i - \theta_j$

Sample in ratios $\sqrt{N_m} : 1 : \ldots : 1$ within stage m.

Problem 1 is dealt with properly in (b); Problem 2 is ignored (Solution A!)

Modern applications: Medical

Aim: Combine phase II and phase III clinical trials by running a single study to select a treatment (e.g., dose level) and compare this treatment with a control.

References:

Thall, Simon and Ellenberg (*Bmka*, 1988)

Schaid, Wieand and Therneau (*Bmka*, 1990)

Proschan, Follmann and Geller (*Statist. in Med.*, 1994)

Stallard and Todd (*Univ. Reading Technical Report*, 1999)

The need to compare with a control treatment changes the problem.

Sequential elimination offers scope for improvement on proposed designs — but there are computational complications.

Modern applications: Computer learning

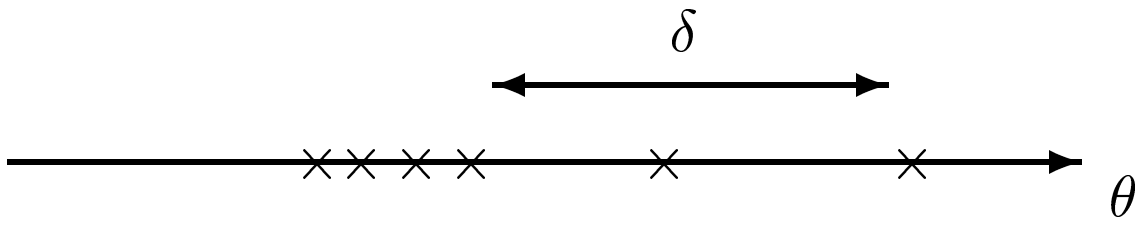
Aim: To select and rank algorithms for their performance on randomly chosen problems.

Experimentation is on a computer, hence any well-defined procedure is easily implemented.

In selecting the best s out of k competitors, k can be large, e.g., 10,000. Clearly, sequential methods are desirable — or, with parallel processing, group sequential methods.

Computer scientists are discovering the selection and ranking literature!

Beyond the indifference zone



What if there is a θ_i within δ of the highest θ_j?

It should be OK to select a population within δ of the best. But can a non-optimal population eliminate the best, then be eliminated itself?

Kao and Lai (*Comm. Statist. Th. Meth.*, 1980) provide a solution, raising the boundary for any pair-wise elimination before the final decision.

This method works for Paulson's procedure with adaptive sampling and can be extended to choosing the best s populations out of k.

Conclusions

- Solutions found in 1982 are still very appropriate.
- Applications for selection procedures are appearing in important research areas.
- Interesting challenges remain.

References

Denne, J.S. (2001). Sample size recalculation using conditional power. *Statistics in Medicine* **20**, 2645–2660.

Jennison, C., Johnstone, I. M. and Turnbull, B. W. (1982) Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In *Statistical Decision Theory and Related Topics III, Vol. 2*, (eds S. S. Gupta and J. Berger) New York: Academic Press **2**, 55–86.

Jennison, C. and Turnbull, B. W. (2000) *Group Sequential Tests with Applications to Clinical Trials*, Chapman and Hall/CRC, 390 pages.

Jennison, C. and Turnbull, B. W. (2001) Group sequential tests with outcome-dependent treatment assignment. *Sequential Analysis*, **20**, 209–234.

Kao, S. C. and Lai, T. L. (1980). Sequential selection procedures based on confidence sequences for normal populations. *Comm. Statist. Theory and Methods*, **A9(16)**, 1657–1676.

Müller, H-H. and Schäfer, H. (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential procedures. *Biometrics* **57**, 886–891.

Paulson, E. (1964) A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Statist.*, **35**, 174–180.

Proschan, M.A., Follmann, D.A and Geller, N.L. (1994). Monitoring multi-armed trials. *Statistics in Medicine*, **13**, 1441–1452.

Robbins, H. and Siegmund, D.O. (1974). Sequential tests involving two populations. *J. Amer. Statist. Assoc.*, **69**, 132–139.

Schaid, D.J., Wieand, S. and Therneau, T.M. (1990). Optimal two-stage screening designs for survival comparisons. *Biometrika*, **77**, 507–513.

Stallard, N. and Todd, S. (1999). Sequential designs for phase III clinical trials incorporating treatment selection. Technical Report 99/1, Department of Applied Statistics, The University of Reading.

Thall, P.F., Simon, R. and Ellenberg, S.S. (1988). Two-stage selection and testing designs for comparative clinical trials. *Biometrika*, **75**, 303–310.