Optimising Group Sequential Designs:

Where Frequentist meets Bayes

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

Improving the Efficiency of Clinical Trials

From Methods to Practice

Bath

June 2025

Plan for this talk

I shall consider :

Frequentist and Bayesian designs for a group sequential trial

What we would like to do

What we can do

Convergence of approaches

Problem formulation

Consider a Phase III clinical trial comparing a new treatment against a control.

We denote the treatment effect by θ.

Examples: θ could be the difference in mean response or, in a time-to-event study, θ could be the log hazard ratio.

We wish to decide whether $\theta > 0$, in which case the new treatment is superior.

The trial will have K analyses.

The information for θ at analysis k will be $\mathcal{I}_k$.

Marginally,

$$\widehat{\theta}_k \sim N(\theta, \mathcal{I}_k^{-1})$$

and the score statistics, $S_k = \widehat{\theta}_k \, \mathcal{I}_k$, have independent increments.

Decision theoretic formulation: Berry & Ho (*Bmcs*, 1988)

Prior distribution $\pi(\theta)$.

Possible decisions

d_1 : Do not pursue drug approval,

d_2 : Pursue drug approval.

Loss function for taking decision d_1 or d_2 when the true value of the treatment effect is θ

$$
\begin{aligned}
L(\theta, d_1) &= 0 \qquad \text{for all } \theta, \\
L(\theta, d_2) &= \begin{cases} -K\theta & \text{if } \theta > 0, \\ L & \text{if } \theta \le 0. \end{cases}
\end{aligned}
$$

Sampling cost: 1 per subject in the trial, N in total, say.

Aim: minimise the expected loss

$$\int \pi(\theta) \int f(x \,|\, \theta) \{L(\theta,\, d(x)) + N(x)\} \, \mathrm{d}x \, \mathrm{d}\theta$$

Decision theoretic formulation: Berry & Ho (*Bmcs*, 1988)

Aim: minimise the expected loss

$$\int \pi(\theta) \int f(x \mid \theta) \{L(\theta, d(x)) + N(x)\} \, \mathrm{d}x \, \mathrm{d}\theta$$

Berry & Ho found the optimal stopping rule and decision rule by dynamic programming.

They presented examples with priors

$$\theta \sim N(-1, 2), \quad \theta \sim N(0, 2), \quad \theta \sim N(1, 2),$$

$$K = 5,000 \quad \text{and} \quad L = 2,000,$$

and showed results for designs with 2 and 3 analyses.

Lorden (*Ann. Statistics*, 1976) had applied the numerical optimisation scheme described by Lai (*Ann. Statistics*, 1973) to solve similar problems in a frequentist setting.

A frequentist problem: Barber & Jennison (*Bmka*, 2002)

Test H_0: $\theta \le 0$ against $\theta > 0$.

Specify type I and type II error rates

$$Pr_{\theta=0}\{\text{Reject } H_0\} \le \alpha, \quad Pr_{\theta=\delta}\{\text{Reject } H_0\} \ge 1-\beta.$$

So, a fixed sample size test requires information

$$\mathcal{I}_{fix} = \{\Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta)\}^2/\delta^2.$$

Aim: in a group sequential test with K analyses, minimise

$$\int f(\theta)\,\mathbb{E}(\mathcal{I}_T)\,\mathrm{d}\theta$$

where $f(\theta)$ is a $N(\delta/2, (\delta/2)^2)$ density and $\mathcal{I}_T$ is the observed information on termination.

Barber & Jennison found optimal stopping and decision rules by dynamic programming.

A frequentist problem: Barber & Jennison (*Bmka*, 2002)

Minimum possible values of $\int f(\theta) E_\theta(\mathcal{I}_T)\, d\theta$, where $f(\theta)$ is the density of a $N(\delta, \delta^2/4)$ distribution, for group sequential tests with K equally sized groups, $\mathcal{I}_{max} = R\,\mathcal{I}_{fix}$, type I error probability $\alpha = 0.025$, power 0.9 at $\theta = \delta$.

Minimum values of $\int f(\theta) E_\theta(\mathcal{I}_T)\, d\theta$, as a percentage of $\mathcal{I}_{fix}$

	R					*Minimum*
K	1.01	1.05	1.1	1.2	1.3	*over R*
2	79.3	74.7	**73.8**	74.8	77.1	73.8 at R=1.11
3	74.8	69.0	67.0	**66.1**	66.6	66.1 at R=1.20
5	71.1	65.1	62.7	60.9	**60.5**	60.5 at R=1.32
10	68.2	62.1	59.5	57.5	**56.7**	56.4 at R=1.46
20	66.8	60.6	58.0	55.8	**54.8**	54.2 at R=1.59

Recommend: $K = 5$, $R = 1.05$ or 1.1.

A frequentist problem: Barber & Jennison (*Bmka*, 2002)

For practical application:

Error spending tests with type I and type II error spending functions of the form

$$f(\mathcal{I}) = \alpha\,(\mathcal{I}/\mathcal{I}_{\max})^{\rho}, \quad g(\mathcal{I}) = \beta\,(\mathcal{I}/\mathcal{I}_{\max})^{\rho}$$

are almost optimal.

Error spending designs adapt to observed information levels, controlling the type I error rate and maintaining efficiency.

Rho-family designs with $\rho = 2$ have a sample size "inflation factor" around $R = 1.1$.

Designs with $\rho = 3$ have an "inflation factor" around $R = 1.05$.

Derivation of optimal tests

The problem is to minimise

$$\int f(\theta)\, E_\theta(\mathcal{I}_T)\, \mathrm{d}\theta,$$

subject to

$$Pr_{\theta=0}\{\text{Reject } H_0\} \le \alpha, \quad Pr_{\theta=\delta}\{\text{Reject } H_0\} \ge 1-\beta. \qquad (1)$$

Following the Lagrangian approach, we use dynamic programming to minimise

$$\int f(\theta)\, E_\theta(\mathcal{I}_T)\, \mathrm{d}\theta + \lambda_1\, Pr_{\theta=0}\{\text{Reject } H_0\} + \lambda_2\, Pr_{\theta=\delta}\{\text{Accept } H_0\}.$$

Then we search for values λ_1 and λ_2 so that (1) is satisfied — and we have the solution to the problem with error rate constraints.

Derivation of optimal tests

In minimising

$$\int f(\theta)\, E_\theta(\mathcal{I}_T)\, \mathrm{d}\theta + \lambda_1\, Pr_{\theta=0}\{\text{Reject } H_0\} + \lambda_2\, Pr_{\theta=\delta}\{\text{Accept } H_0\},$$

we have solved a Bayes decision problem with prior

$$\begin{array}{ll} \theta = 0 & \text{with probability } 1/3 \\ \theta = \delta & \text{with probability } 1/3 \\ \theta \sim N(\delta/2, \delta^2/4) & \text{with probability } 1/3 \end{array}$$

and a cost of sampling $\mathcal{I}_T$ when $\theta \sim N(\delta/2, \delta^2/4)$.

This is a Bayes optimal procedure — but for a rather odd looking problem.

Back to Bayesian proposals

Spiegelhalter, Freedman & Parmar (1994) proposed the following sequential procedure. Given a prior $\pi(\theta)$,

At each analysis $k = 1, \ldots, K-1$

if $Pr\{\theta > 0 \mid \textit{data}\} > 1 - \epsilon$	stop, declare $\theta > 0$
if $Pr\{\theta < 0 \mid \textit{data}\} > 1 - \epsilon$	stop, declare $\theta \leq 0$
otherwise	continue to group $k + 1$,

after group K

if $Pr\{\theta > 0 \mid \textit{data}\} > 1 - \epsilon$	stop, declare $\theta > 0$
otherwise	stop, do not declare $\theta > 0$.

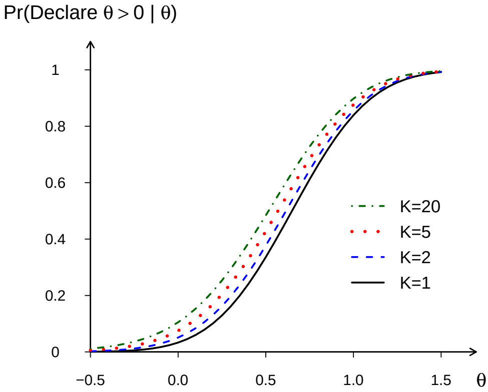
The trial stops early if the $1 - 2\epsilon$ credible interval for θ does not contain zero.

The credible interval is not affected by the fact that earlier analyses have been conducted.

Bayesian proposals: Spiegelhalter et al. (1994)

However, the frequentist properties of the Bayes procedure **are** affected by the number of analyses conducted.

Probability of declaring $\theta > 0$ for procedures with prior $\theta \sim N(0.5, 0.5)$, $\epsilon = 0.025$, and 1, 2, 5 and 20 analyses:



Bayesian proposals: Calibrated procedures

Spiegelhalter et al. (1994) proposed use of a "handicap prior", chosen so that the procedure has a particular type I error rate.

If the number of analyses is K and maximum information is $\mathcal{I}_K$, the handicap prior is

$$\theta \sim N(0, (h\mathcal{I}_K)^{-1}).$$

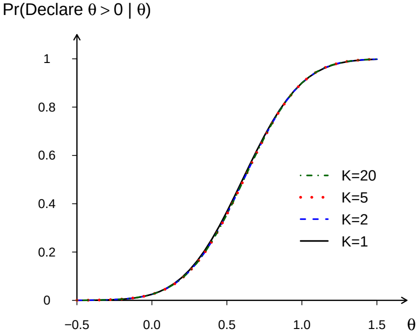
The "handicap" h depends on the number of analyses. Here are values for $\alpha = 0.025$.

Number of analyses K	*Handicap* h
1	0.000
2	0.163
5	0.271
20	0.382

Bayesian proposals: Calibrated procedures

We can choose $\mathcal{I}_K$ so that the procedure with a handicap prior has a specific power if the treatment effect is $\theta = \delta$.

Power functions of designs with $Pr\{\text{Reject } H_0 \,|\, \theta = 1\} = 0.9$



Note: When power curves are matched at $\theta = 0$ and $\theta = 1$, they are just about indistinguishable everywhere.

Bayesian proposals: Ventz & Trippa (2015)

Ventz & Trippa proposed optimising and calibrating at the same time. Their problem formulation has:

Possible decisions

$$\begin{aligned} d_1 &: \quad \text{Do not pursue drug approval,} \\ d_2 &: \quad \text{Pursue drug approval.} \end{aligned}$$

A “gain function” or “utility” comprising

$$\begin{aligned} G(\theta, d_1) &= 0 \qquad\qquad \text{for all } \theta, \\ G(\theta, d_2) &= \begin{cases} K(\theta) & \text{if } \theta > 0, \\ -L(\theta) & \text{if } \theta \leq 0, \end{cases} \end{aligned}$$

plus a term

$$B(\theta, d_i, T)$$

denoting the additional benefit from reaching decision d_i at analysis T minus the cost of treating patients in the trial.

Bayesian proposals: V&T Calibrated Bayes procedure

Assuming a prior distribution $\pi(\theta)$, Ventz & Trippa seek to minimise the expected gain

$$-\int_{-\infty}^{0} L(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta + \int_{0}^{\infty} K(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta + \int_{-\infty}^{\infty} E\{B(\theta, d_i, T)\}\, \pi(\theta)\, d\theta.$$

subject to error rate constraints

$$Pr(D = d_2 \,|\, \theta = 0) \;=\; \alpha \quad \text{and} \quad Pr(D = d_1 \,|\, \theta = \delta) \;=\; \beta$$

for a specified value of δ.

Note that type I and type II errors feature twice, in different guises.

Bayesian proposals: V&T Calibrated Bayes procedure

On Slide 14, we saw an example of how, to a high degree of accuracy, power belong to a two parameter family.

Thus, the constraints

$$Pr(D = d_2 \,|\, \theta = 0) \;=\; \alpha \quad \text{and} \quad Pr(D = d_1 \,|\, \theta = \delta) \;=\; \beta$$

are essentially equivalent to

$$\int_{-\infty}^{0} L(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta \;=\; p_1$$

and

$$\int_{0}^{\infty} K(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta \;=\; p_2$$

for certain values of p_1 and p_2.

Bayesian proposals: V&T Calibrated Bayes procedure

Thus, when Ventz & Trippa minimise

$$-\int_{-\infty}^{0} L(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta \;+\; \int_{0}^{\infty} K(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta$$

$$+\; \int_{-\infty}^{\infty} E\{B(\theta, d_i, T)\}\, \pi(\theta)\, d\theta$$

subject to

$$Pr(D = d_2 \,|\, \theta = 0) \;=\; \alpha \quad \text{and} \quad Pr(D = d_1 \,|\, \theta = \delta) \;=\; \beta.$$

they are effectively minimising

$$\int_{-\infty}^{\infty} E\{B(\theta, d_i, T)\}\, \pi(\theta)\, d\theta.$$

subject to

$$\int_{-\infty}^{0} L(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta = p_1 \;\text{ and }\; \int_{0}^{\infty} K(\theta)\, Pr(D = d_2 \,|\, \theta)\, \pi(\theta)\, d\theta = p_2.$$

Bayesian proposals: V&T Calibrated Bayes procedure

Given the equivalence of the two types of constraint on the power function, they are also minimising

$$\int_{-\infty}^{\infty} E\{B(\theta, d_i, T)\}\, \pi(\theta)\, d\theta.$$

subject to

$$Pr(D = d_2 \,|\, \theta = 0) \;=\; \alpha \quad \text{and} \quad Pr(D = d_1 \,|\, \theta = \delta) \;=\; \beta.$$

This is exactly the type of problem that Jennison and others have tackled from a frequentist perspective.

A remark on the frequentist designs we saw earlier

I noted that the designs produced by Barber & Jennison were Bayes procedures for rather strange looking priors.

However, we can replace the constraint

$$Pr(D = d_2 \,|\, \theta = 0) \;=\; \alpha \quad \text{and} \quad Pr(D = d_1 \,|\, \theta = \delta) \;=\; \beta$$

by an (almost) equivalent constraint of the form

$$\int_{-\infty}^{0} Pr(D = d_2 \,|\, \theta)\, f(\theta)\, d\theta = p_1 \;\; \text{and} \;\; \int_{0}^{\infty} Pr(D = d_2 \,|\, \theta)\, f(\theta)\, d\theta = p_2,$$

where $f(\theta)$ is a $N(\delta/2,\, (\delta/2)^2)$ density.

Then, we will obtain essentially the same optimal design by solving a more reasonable looking Bayesian problem.

CONVERGENCE!

Conclusions

Theoretical underpinnings

The convergence between frequentist and calibrated Bayes procedures demonstrates the "complete class theorems" of Brown, Cohen & Strawderman (*Ann. Statistics*, 1980) which state

$$\{\text{The set of admissible frequentist procedures}\}$$

$$= \{\text{The set of Bayes optimal procedures}\}.$$

Practical consequences

Both schools can learn from each other — we are (or should be) solving the same problems.