Bayesian randomized adaptive designs with a compound utility function

Alessandra Giovagnoli
University of Bologna, Italy,
and
Isabella Verdinelli
Carnegie Mellon University, USA

Design of Experiments: New Challenges
CIRM – Marseille 2018

Outline

1. Optimal experiments for treatment comparison

   Why Bayesian? Why Randomized? Why Adaptive?

2. The utility-based Bayes adaptive designs
   - Choice of utility

   _____

3. An *A*-optimal Bayes adaptive design for binary data
   - Convergence to optimal target
   - Discussion

4. Bayesian doubly-adaptive methods
   - The *Bayesian Randomized Adaptive Compound* (BRAC)
   - The *Bayesian Randomized Adaptive Compound Efficient* (BRACE) design
   - Convergence to optimal target

5. Some simulations to compare BRAC and BRACE

Why Bayesian?

Optimal experiments for treatment comparison

Nearly all desirable treatment allocations (= *targets*) depend on unknown parameters of model

Example :

- Two treatments T_1, T_2
- binary responses, with success probabilities p_1, p_2
- Minimizing the variance of estimated difference p_1 - p_2 gives **Neyman**'s allocation proportion

$$\frac{n_1}{n} = \frac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}$$

With 50-50 allocation

If $p_1 = 0.1$ and $p_2 = 0.4$, loss of information = 5.8%

If $p_1 = 0.2$ and $p_2 = 0.3$, loss of information = 4.6 $^{o}/_{oo}$

Why Bayesian?

Need to use prior information on treatment effects <u>for designing the experiment</u>

- *best guess*
- *Bayesian* – prior distribution on parameters

(very popular **also among non-Bayesians!**)

Wish to **use prior distribution AT THE PLANNING STAGE, but not necessarily for inference**

Bayarri & Berger *Statistical Science,* 2004;

Ventz, *et al. Applied Stochastic Models in Business and Industry*, 2017

Why Randomized?

Randomization not essential in Bayesian statistics, **BUT** now widely accepted
(Ball, Smith and Verdinelli *JSPI*, 1993)

1. Randomization to fight accidental/selection bias
2. Present trend:

Bayesian approach to design strongly biased towards clinical studies

S. Berry, Carlin, Lee and Muller, (2011)

Bayesian Adaptive Methods for Clinical Trials

FDA guidelines (2010)

In clinical trials randomization is a must

Why Adaptive?

Designing whole experiment on the basis of the initial prior can be very inefficient

More efficient to update prior belief step-by-step in the light of the accrued data

Adaptive design in a Bayesian framework: how?

Clearly

Taking the prior expectation of an optimal target may turn out to be strongly sub-optimal

(*falling between two stools!* ☺)

Several possible Bayesian approaches: "probability only", Bayesian biased coins....

Our choice: **The Decision-theoretic Approach** (Lindley, 1957)

Start with

- a prior on parameters and a suitable utility function

At each step

- update the prior
- calculate the expected utility according to updated prior
- maximize the expected utility (= *temporary targets*)
- use a randomization rule that is a function of the temporary target

Choice of utility (or loss)

In Chaloner and Larntz (*JSPI*, 1989):

Utility as a function of Fisher's information matrix $\mathcal{M}$

Information criteria

- Trace of inverse Fisher's info — **A-optimality**
- Determinant of Fisher's info — **D-optimality**

have Bayesian interpretations

"Ethical" utilities (in clinical trials)

- Expected number of subjects allocated to better treatment/s
- Expected number of successes

Compound utilities (*Trade-offs of Exploration v. Exploitation*)

Weighted combinations of above (with weights possibly depending on the parameters)

see also Asya Metelkina & Luc Pronzato *Information-regret compromise in covariate-adaptive treatment allocation*

An *A*-optimal Bayes adaptive design for binary data

(in Applied Stochastic Models in Business and Industry, 2017)

Two treatments $T_1 \quad T_2$

Model is binary success probabilities $p_1 \quad p_2$

Utility is A-optimality: want to minimize

$$E_{(p_1,p_2)}\left[trace\begin{pmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{pmatrix}^{-1} \right]$$

$\Rightarrow$ **A-optimal Bayes design**:

Proportion of allocations to T_1

$$\pi_B = \frac{\sqrt{E_{p_1}[p_1q_1]}}{\sqrt{E_{p_1}[p_1q_1]} + \sqrt{E_{p_2}[p_2q_2]}}$$

NOT the same as expected Neyman's target

$$E_{(p_1,p_2)}\left[\frac{\sqrt{p_1q_1}}{\sqrt{p_1q_1} + \sqrt{p_2q_2}} \right]$$

An *A*-optimal Bayesian adaptive design for binary data

Priors are independent *Betas* $Beta(a_j,b_j)$ $j=1,2$

$\Longrightarrow$ Posteriors are *Betas*

At each step $k \geq 1$

N_{jk} = number of assignments to T_j

S_{jk} = number of successes of T_j

A-optimal allocation of T_1 *according to the updated information (temporary target) is*

$$\mathbf{T}(k) = \frac{G_{1k}}{G_{1k} + G_{2k}}$$

where $$G_{jk} = \frac{\sqrt{(a_j + S_{jk})(b_j + N_{jk} - S_{jk})}}{\sqrt{(a_j + b_j + N_{jk})(a_j + b_j + N_{jk} + 1)}}$$

Randomization:

At step k, choose T_1 with probability equal to temporary target

A-optimal Bayesian adaptive design

The treatment allocation proportion N_{1n}/n converges almost surely to Neyman's target π_N for all *Beta* priors and all p_1 p_2

This A-optimal Bayes adaptive design is Bayesian analogue of frequentist **Sequential Maximum Likelihood** design targeting Neyman's allocation

The **Sequential Maximum Likelihood (**SML) design *for a target* $\pi_0(\theta)$

n_0 observations for initial estimate of parameters

at each step i ($> n_0$)

- parameter θ estimated by ML utilizing all the data up to k
- the unknown target $\pi_0(\theta)$ is estimated
- treatments are randomized with probability given by the current estimate of the target regardless of whether present assignments are too few or too many

- Almost sure convergence of proportion of allocations N_{1n}/n to target
- ML estimators are consistent and asimptotically normal
- Asymptotic normality of N_{1n}/n

BUT Slow convergence

An *A*-optimal Bayesian adaptive design for binary data

Some Beta priors $Beta(a_j, b_j)$

$p_1^* = 0.1$

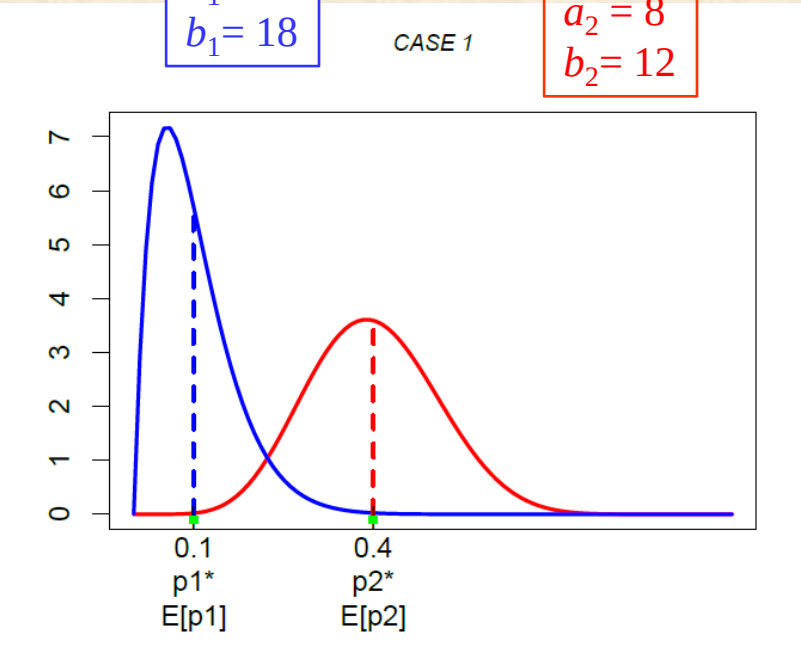
$p_2^* = 0.4$

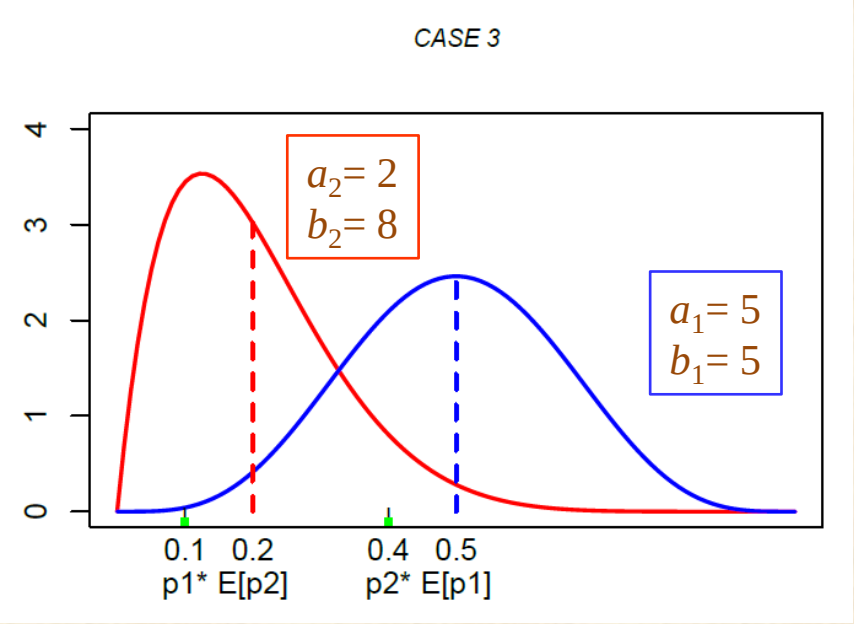
CASE 2

$a_1 = 1$, $b_1 = 1$ $\quad$ $a_2 = 1$, $b_2 = 1$

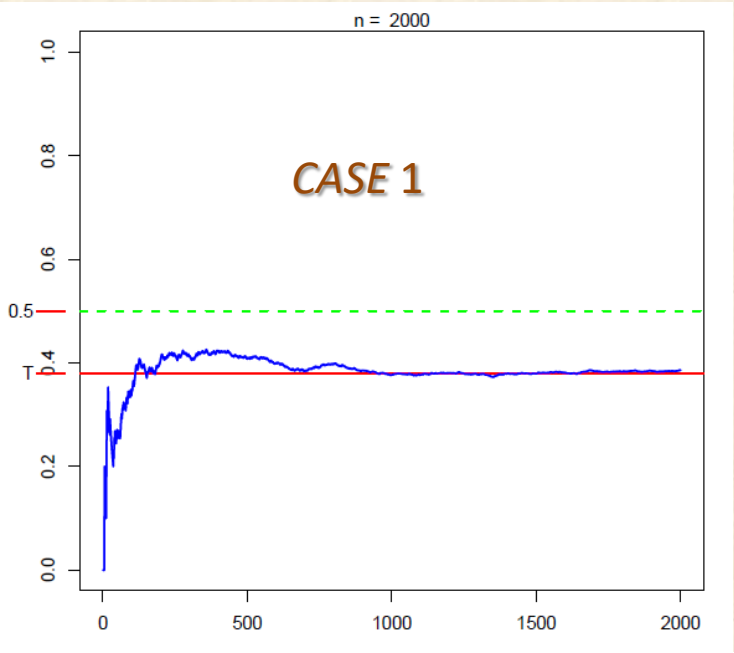
Uniform in [0,1]

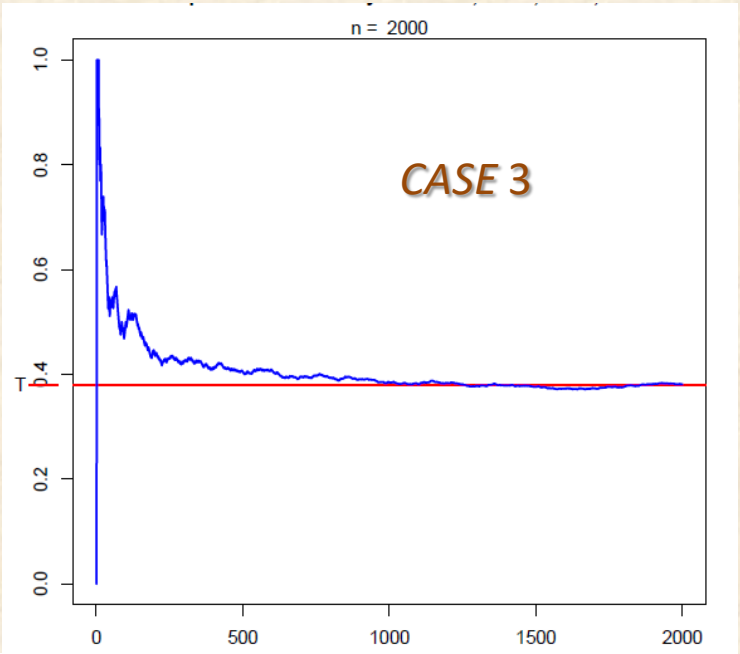
$a_1 = 2$, $b_1 = 18$ $\quad$ $a_2 = 8$, $b_2 = 12$

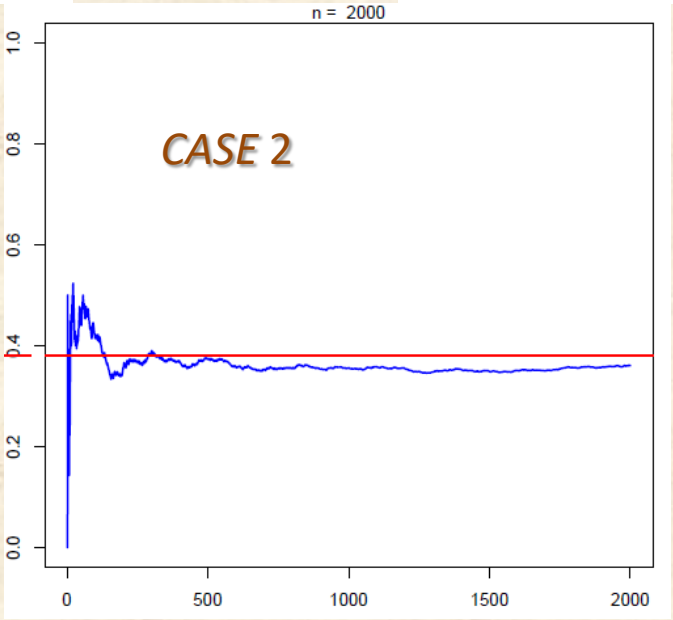




Convergence of Bayes A-optimal design to Neyman's target







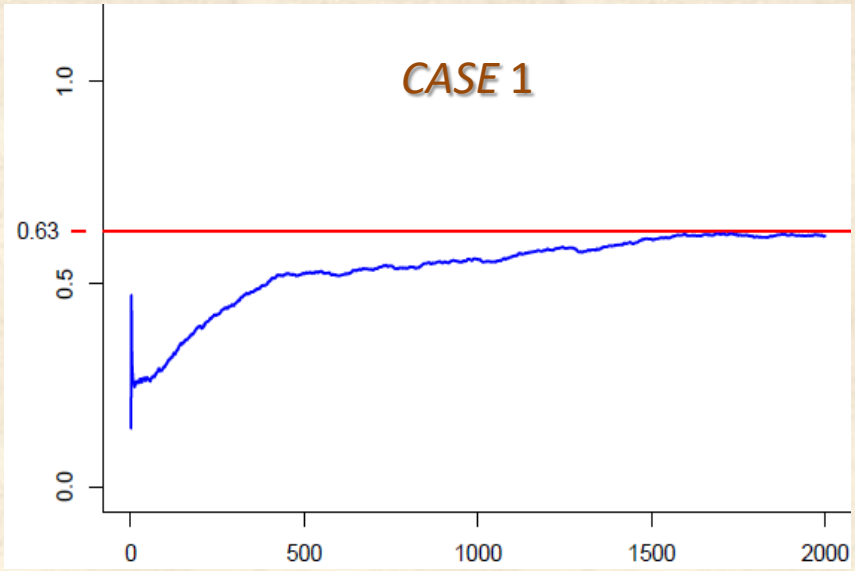
p_1*= 0.1

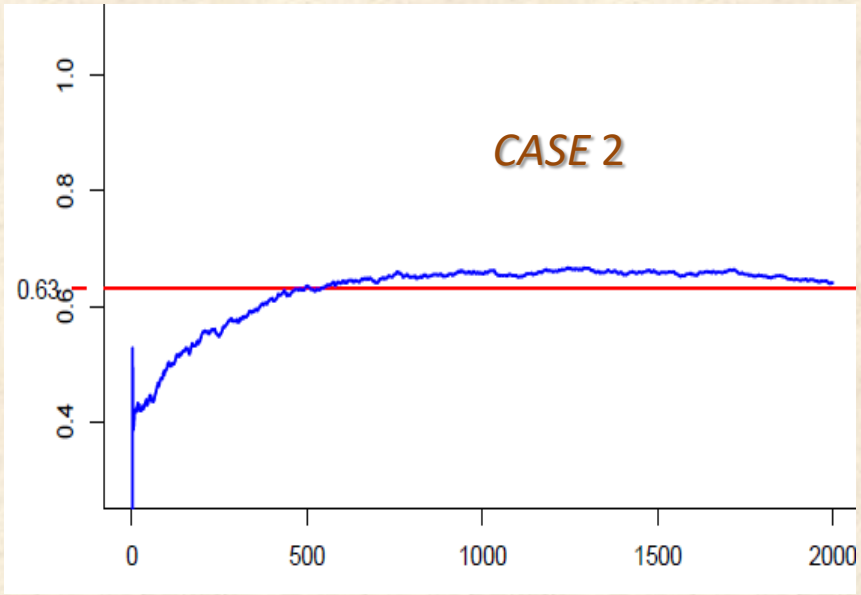
p_2*= 0.4

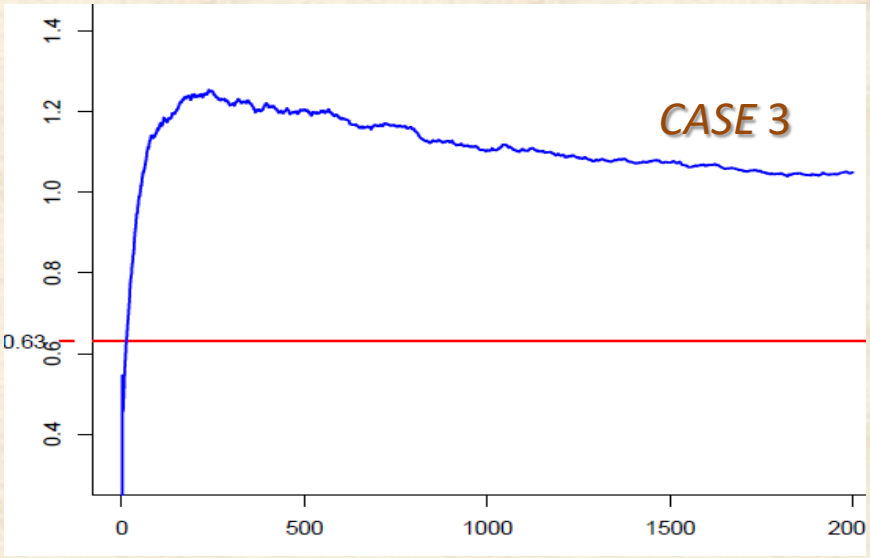
Neyman's target = 0.382

Asymptotic variance of A-optimal Bayes design

p_1*= 0.1, p_2*= 0.4 ; No. of replications = 1000







The efficiency of a randomized design is a decreasing function of the variability of the design

(**Hu & Rosenberger**, JASA, 2003)

Asymptotic variance

of the SML design = 0.76

To get faster convergence (less variability) can we import frequentist improvement methods into Bayesian designs?

Compound utilities

Ethical/Utilitarian criterion

$\mathcal{E}$ = Expected proportion of units allocated to better treatment

Information criterion

Φ = determinant, square root of det, trace of inverse normalized Fisher information $\mathcal{M}$

Essential requirement: $\underline{\Phi}$ is Φ re-scaled to range in [0,1]

$$\mathcal{U} = \omega\ \mathcal{E} + (1 - \omega)\underline{\Phi} \qquad 0 \le \omega \le 1$$

$$\mathcal{U} = \mathcal{U}(\pi; \omega; p_1^*, p_2^*) \qquad \pi = n_1/n$$

At stage k, we maximize expected utility wrt posterior up to k-1

"TEMPORARY TARGETS"

$$\mathbb{T}_{\mathcal{U}}(k) = \arg\max_{\pi} E(\mathcal{U} \mid \mathcal{F}_{k-1})$$

Examples

1. $\Phi = \det \mathcal{M}$

$$\mathbb{T}_D(k) = \frac{1}{2} + \frac{1}{8}\frac{\omega}{1-\omega}\left(\mathbb{P}(p_1 > p_2|\mathcal{F}_{k-1}) - \mathbb{P}(p_1 < p_2|\mathcal{F}_{k-1})\right)$$

$\omega < 4/5$

With Beta priors

$$\mathbb{T}_D(k) = \frac{1}{2} + \frac{1}{8}\frac{\omega}{1-\omega}\left(2\,g\,(a_1^{(k)}, b_1^{(k)}, a_2^{(k)}, b_2^{(k)}) - 1\right)$$

where

$$g(a_1, b_1, a_2, b_2) = \int_0^1 \frac{x^{a_1-1}(1-x)^{b_1-1}}{B(a_1, b_1)} I_x(a_2, b_2)\, dx$$

and $I_x(a_2, b_2)$ is the CDF of Beta(a_2, b_2)

2. $\Phi = \text{SQRT}(\det \mathcal{M})$

Beta priors

$$\mathbb{T}_{\sqrt{D}}(k) = \frac{1}{2} + \frac{\omega\left(2g(a_1^{(k)}, b_1^{(k)}, a_2^{(k)}, b_2^{(k)}) - 1\right)}{\sqrt{4(1-\omega)^2 + \omega^2\left(2g\,(a_1^{(k)}, b_1^{(k)}, a_2^{(k)}, b_2^{(k)}) - 1\right)^2}}.$$

3. $\Phi = \text{trace}(\mathcal{M}^{-1})$

$\omega < 1/2$

Simply adaptive randomization rule

Randomization :

at stage k $\mathbb{P}(T_1) = \mathbf{T}_U(\boldsymbol{k})$ Optimal allocation given the data

or $\mathbb{P}(T_1) = \varphi(\mathbf{T}_U(\boldsymbol{k}))$

1. $\varphi(y) \nearrow$ in y
2. $\varphi(y) = 1 - \varphi(1-y)$ for all y

When utility is <u>compound</u>,

Bayesian Randomized Adaptive Compound (*BRAC*)

design

Doubly-adaptive designs

In frequentist context:

Doubly-adaptive randomization:

takes into account both past allocations and target

The probability of assigning T_1 to unit n+1 is based on a measure of "dissimilarity" between the current allocation of T_1 and the current estimate of the target

- **Doubly-adaptive Biased Coin Design**
  (Eisele, *JSPI*, 1994; Hu & Zhang, *Annals*, 2004)
- **Efficient Randomized-Adaptive DEsign** (ERADE)
  (Hu, Zhang & He, *Annals,* 2009)
- **Reinforced Doubly-Adaptive BCD**
  (Baldi-Antognini & Zagoraiou, *Annals*, 2012)

(**D**oubly-adaptive **BCD** and **ERADE** are special cases)

Doubly-adaptive randomization rule

$$\psi: (0,1)^2 \rightarrow [0,1]$$

(C) Conditions for $\psi(x,y)$

1. $\psi(x, y) \searrow$ in x
2. $\psi(x, y) \nearrow$ in y
3. $\psi(x, x) = x$ for all x
4. $\psi(x, y) = 1 - \psi(1-x, 1-y)$ for all x, y

x = present allocation proportion

y = current temporary target

ψ = some measure of dissimilarity

Randomization : Choose T_1 with probability = $\psi\left(N_{1k}/k, \mathbb{T}_{\mathcal{U}}(k)\right)$

If the allocation proportion N_{1k}/k is smaller than the temporary target $\mathbb{T}_{\mathcal{U}}(k)$ then the allocation probability will be greater than the temporary target and viceversa. This will drive allocations to the true target if $\mathbb{T}_{\mathcal{U}}(k) \rightarrow \mathbf{T}^*$

Convergence of Bayesian doubly-adaptive designs

Theorem

(after Baldi Antognini & Zagoraiou, *Bernoulli*, 2015)

If

i) the utility function $\mathcal{U}$ and independent priors for p_1 and p_2 are such that $0 < \mathbb{T}_{\mathcal{U}}(k) < 1$ and

ii) the Bayesian randomized doubly-adaptive design is defined by ψ satisfying **(C)**,

then

- **The temporary targets converge to the desired target almost surely**
- **The allocation proportion converges to the desired target almost surely**

ERADE

(**Hu, Zhang & He**, *Annals,* 2009 – inspired by **Efron**'s **Biased Coin Design**, 1971)

Frequentist

$$\mathbb{P}(T_1) = \begin{cases} \rho y, & \text{if } x > y, \\ y, & \text{if } x = y, \\ 1 - \rho(1-y), & \text{if } x < y, \end{cases}$$

- x = present allocation proportion
- y = current estimate of the target

$\rho \in [0; 1)$ a randomization parameter : for $\rho = 0$ ERADE becomes design by **Robbins et al.** (1967), for $\rho \to 1$ allocations are more randomized and ERADE tends to the **SML design**

- Almost sure convergence and asymptotic normality hold.
- Under differentiability conditions of the target, the asymptotic variance is a minimum.

ERADE is asymptotically best

BRACE design

(Bayesian Randomized Adaptive Compound Efficient)

Randomization rule = “Bayesian ERADE”

$$\mathbb{P}(T_1) = \begin{cases} \rho\mathbb{T}(k) & \text{if } n_{1k}/k > \mathbb{T}(k) \\ \mathbb{T}(k) & \text{if } n_{1k}/k = \mathbb{T}(k) \\ 1-\rho\ (1-\mathbb{T}(k)) & \text{if } n_{1k}/k < \mathbb{T}(k) \end{cases}$$

Satisfies Condition **(C)** → **Convergence theorem applies**

Comparisons of Bayesian adaptive and doubly adaptive designs

Xiao, Liu & Hu (2017, *Science China: Mathematics*) also study Bayesian adaptive designs

- No utility function.
- Randomization rule is a function of $\mathbb{P}(p_1 > p_2|\mathcal{F}_k)$.
- They have a threshold value for the randomization probability.
- They compare *BAR* (simply adaptive) with *BDBCD* (doubly-adaptive, based on **Hu and Zhang**, 2004)
- They prove convergence of BDBCD to a fixed value which depends on the threshold, and the asymptotic normality of BDBCD.

Comparing BRAC and BRACE designs

Example 1

$$\mathbb{T}_D(k) = \frac{1}{2} + \frac{1}{8}\frac{\omega}{1-\omega}\left(2\,g\,(a_1^{(k)}, b_1^{(k)}, a_2^{(k)}, b_2^{(k)}) - 1\right)$$

p_1*= 0.4

p_2*= 0.5

ω = 0.5, ρ = 0.5

PRIORS

a_1 = 40, a_2 = 60

b_1 = 50, b_2 = 50

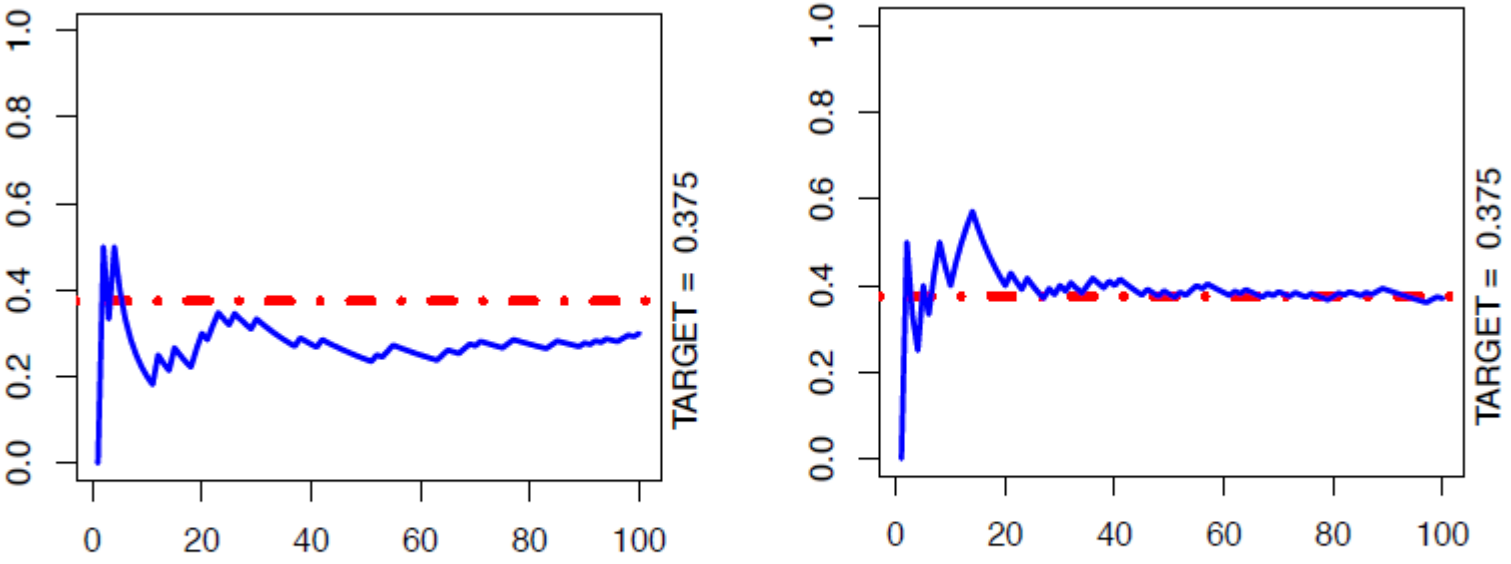


Figure 5. Single run of the algorithm. Left: BRAC. Right: BRACE.

Comparing BRAC and BRACE designs

Example 1

$$\mathbb{T}_D(k) = \frac{1}{2} + \frac{1}{8}\frac{\omega}{1-\omega}\left(2\,g\left(a_1^{(k)}, b_1^{(k)}, a_2^{(k)}, b_2^{(k)}\right) - 1\right)$$

$p_1^* = 0.4$
$p_2^* = 0.5$

$\omega = 0.5$

$a_1 = 5, a_2 = 5$
$b_1 = 4, b_2 = 5$

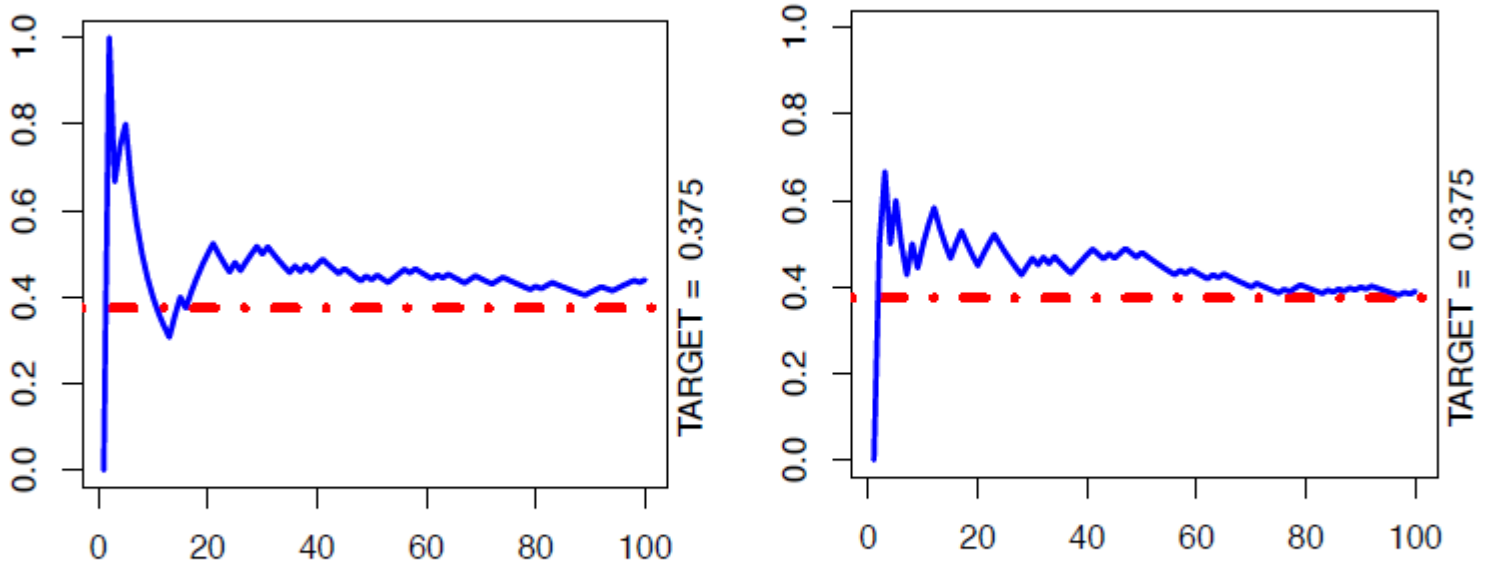


Figure 6. Single run of the algorithm. Left: BRAC. Right: BRACE.

Comparing BRAC and BRACE designs

Example 1 | N = 1000 replications | $\omega = 0.5$ | "Bad" prior

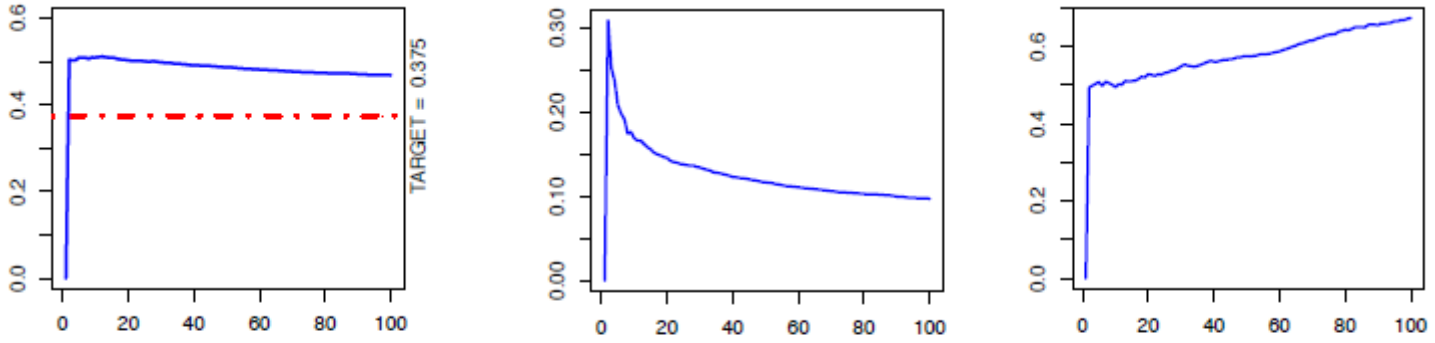


Figure 3. BRAC design. $p_1^* = 0.4$, $p_2^* = 0.5$.
Left: $\mathbb{E}\left[n_{1n}/n\right]$. Center: $\mathbb{E}\left[|n_{1n}/n - \mathbb{T}^*|\right]$. Right: AVAR= $\sqrt{k}\,SD[n_{1k}/k]$; $k = 1, \ldots, n$.

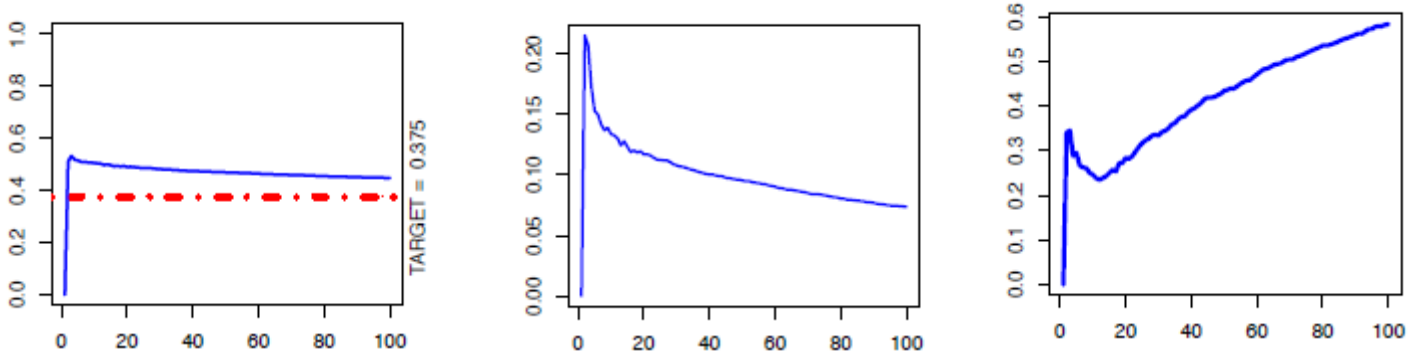


Figure 4. BRACE design. $p_1^* = 0.4$, $p_2^* = 0.5$.
Left: $\mathbb{E}\left[n_{1n}/n\right]$. Center: $\mathbb{E}\left[|n_{1n}/n - \mathbb{T}^*|\right]$. Right: AVAR= $\sqrt{k}\,SD[n_{1k}/k]$; $k = 1, \ldots, n$.

Comparing BRAC and BRACE designs

N = 1000 replications

ω = 0.5

Choice of Beta priors

	a_1	b_1	a_2	b_2
(A)	40	60	50	50
(B)	4	6	5	5
(C)	5	5	4	5
(D)	50	50	40	50

$p_1^* = 0.4,\ p_2^* = 0.5,$ n=100. $\mathbb{T}^* = 0.375.$

	BRAC		BRACE	
	$\mathbb{E}[n_{1n}/n]$	AVAR	$\mathbb{E}[n_{1n}/n]$	AVAR
PRIOR (A)	0.394	0.498	0.393	0.184
PRIOR (B)	0.438	0.648	0.431	0.507
PRIOR (C)	0.472	0.684	0.448	0.561
PRIOR (D)	0.533	0.590	0.509	0.479

(A) is “very good” prior, (B) is “good”, (C) and (D) are “bad”

Average Variability $\mathrm{AVAR} = \sqrt{n}\ SD[n_{1n}/n]$

Comparing BRAC and BRACE designs

N = 1000 replications

ω = 0.5

	a_1	b_1	a_2	b_2
(E)	16.3	65.2	20	13.3
(F)	1.3	5.2	5.1	3.4
(G)	5.1	3.4	1.3	5.2
(H)	20	13. 3	16.3	65.2

$p_1^* = 0.2$, $p_2^* = 0.6$, n=100. $\mathbb{T}^* = 0.375$.

	BRAC		BRACE	
	$\mathbb{E}[n_{1n}/n]$	AVAR	$\mathbb{E}[n_{1n}/n]$	AVAR
PRIOR (E)	0.379	0.484	0.378	0.071
PRIOR (F)	0.378	0.480	0.379	0.069
PRIOR (G)	0.430	0.542	0.381	0.115
PRIOR (H)	0.596	0.536	0.531	0.462

Table 5. *Summary of results for BRAC and BRACE designs. Larger Difference between p_1^* and p_2^*.*

(E) is “very good” prior, (F) is “good”, (G) and (H) are “bad”

Conclusions

Simulations show:

- Convergence is better when the success probabilities are farther apart.
- If the initial perception regarding the treatment effects is fundamentally correct, then a small number of units *n is* sufficient to achieve convergence.
- The prior precision does not seem to affect the design much.
- The ERADE modifications seem improve the speed of convergence to the optimal target in the Bayesian approach too

TO DO

- Our simulations only carried out for one compound utility: ample scope for further investigation.
- The asymptotic normality of the design in the Bayesian context still needs to be investigated
- COVARIATES!!!

THE END

with thanks to the organizers for inviting me

References

BOOKS

- **ROSENBERGER, W.F. and LACHIN, J.M. (2002)** ***Randomization in Clinical Trials, New York , Wiley & Sons.***
- **HU, F. and ROSENBERGER, W.F. (2006)** ***The Theory of Response-Adaptive Randomization in Clinical Trials. New York, Wiley & Sons.***
- **ATKINSON, A. C. and BISWAS, A. (2014)** ***Randomised Response-Adaptive Designs in Clinical Trials. Boca Raton: Chapman & Hall/CRC Press***
- **BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2015)** ***Adaptive Designs for Sequential Treatment Allocation Boca Raton: Chapman & Hall/CRC Press***

- A. Baldi-Antognini and A. Giovagnoli (2010). *Compound optimal allocation for individual and collective ethics in binary clinical trials*, **Biometrika**, 97: 935-946.
- A. Baldi-Antognini and M. Zagoraiou (2015) *On the almost sure convergence of adaptive allocation procedures*. **Bernoulli**, 21(2):881908
- A. Giovagnoli (2017). *Bayesian optimal experimental designs for binary responses in an adaptive framework.* **Applied Stochastic Models in Business and Industry** 33 260–268**.**
- F. Hu, L.X. Zhang and X. He (2009). *Efficient randomized-adaptive designs*, **The Annals of Statistics** 37: 2543-2560.