Bandit Procedures for Designing Patient-Centric Clinical Trials

Peter Jacko*

Workshop on Restless Bandits, Grenoble
20 November 2023

*Berry Consultants & Lancaster University Management School, UK

Restless Bandits in My Research

- 38/77 publications in my Google Scholar profile
  - Scheduling in (time-varying) wireless networks
  - Congestion avoidance in the Internet
  - Scheduling in systems with abandonments
  - Dynamic promotion in marketing
  - Allocation of funding to prevent company defaults
  - Dynamic allocation of failing assets
  - Adaptive treatment allocation in clinical trials
  - Deadlines, finite horizon
  - Arms with absorbing states, no recurrent states
  - New arms arrivals, control arms
  - Heterogeneous resource requirements

Multi-Armed Bandit Problem

- A decision-making problem addressing the omnipresent trade-off between learning and earning
  - A heuristic long known to gamblers: stay on a success, switch on a failure
  - Formulated by scientists in 1930s/40s/50s
  - Classic problem in applied probability, addressed by methods from statistics, OR, machine learning, etc.
  - "Sir, the multi-armed bandit problem is not of such a nature that it can be solved"
  - Variants motivated by applications in health care, economics, marketing, telecommunications, etc.

Multi-Armed Bandit Problem

em addressing the omnipresent
ng and earning

to gamblers: stay on a
ilure

applications in health care

Springer Series in Supply Chain Management

Xi Chen
Stefanus Jasin
Cong Shi *Editors*

The Elements of Joint Learning and Optimization in Operations Management

Randomised Controlled Trials

- The gold standard design for 2 arms:
  - ▷ equal (50% vs 50%) fixed randomisation (EFR)
  - ▷ in use since 1948 (advocated since Hill 1937)
- Its main goal is to learn about intervention effectiveness with a view to prioritise after-trial subjects
  - ▷ the intervention effect estimate is unbiased and unaffected by time trends (if equal on both arms)
  - ▷ if approved, future subjects are, say, 95% confident that the novel intervention is better than the control
- A half of trial subjects gets the inferior intervention

Randomised Controlled Trials

- Frequentist statistical testing based on EFR is a widespread state-of-the-art approach in the design of experiments, under different names:
  - ▷ randomised controlled trial in biostatistics
  - ▷ between-group design in social sciences
  - ▷ A/B testing in digital marketing
- However, "controlled" means to compare the novel intervention against a control intervention under the circumstances of the same trial
  - ▷ departing from EFR may improve on other objectives than estimation

Bayesian Randomised Controlled Trials

"...there can be no objection to the use of data, however meagre, as a guide to action required before more can be collected ... Indeed, the fact that such objection can never be eliminated entirely—no matter how great the number of observations—suggested the possible value of seeking other modes of operation than that of taking a large number of observations before analysis or any attempt to direct our course... This would be important in cases where either the rate of accumulation of data is slow or the individuals treated are valuable, or both."

Bayesian Randomised Controlled Trials

- Proposed in Thompson (1933) (pre-dates Hill 1937)
- The goal is to provide higher benefit to both in-trial subjects and after-trial subjects
  - ▷ as opposed to the EFR's learning goal of reliable intervention effect estimation
- It is done by deciding the allocation, i.e., the randomisation probabilities for every subject (or for a group of subjects)
  - ▷ response-adaptive: decisions based on the responses accumulated so far, using Bayesian updating

Patient-Centric Clinical Trials

- Focusing on patient benefit
- Important because healing patients is the ultimate goal of new treatment development
  - ▷ optimally solving learning/healing trade-off
  - ▷ both learning and healing takes place during the trial

Patient-Centric Clinical Trials

- Focusing on patient benefit

Statistical Science
2015, Vol. 30, No. 2, 199–215
DOI: 10.1214/14-STS504

Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges

Sofía S. Villar, Jack Bowden and James Wason

Abstract. *Multi-armed bandit* problems (MABPs) are a special type of optimal control problem well suited to model resource allocation under uncertainty in a wide variety of contexts. Since the first publication of the optimal

Patient-Centric Clinical Trials

EU-PEARL

EU PATIENT-CENTRIC
CLINICAL TRIAL PLATFORMS

Make clinical trials more efficient so that new drugs can be developed faster with proportionally fewer patients receiving placebo or the standard care.

SHAPING THE FUTURE OF CLINICAL TRIALS

Bernoulli Bandit Model

- Finite horizon: T sequentially arriving subjects $t = 1, ..., T$
- Two-armed: intervention $k \in \{C, D\}$ for each subject
- Binary endpoints: success (1) or failure (0)
- θ_k is the unknown success probability of intervention k
- Subject t's response from intervention k is

$$X_{k,t} \sim \text{Bernoulli}(\theta_k)$$

Bayesian Approach

- Replace each θ_k by random variable $\widetilde{\theta}_k$, whose distribution is updated over the trial
- Prior Distribution: $\widetilde{\theta}_k \sim \text{Beta}(s_k(0),\ f_k(0))$ where we take $s_k(0) = f_k(0) = 1$ (uninformative; uniform distr.)
- Posterior Distribution: After observing s_k (f_k) successes (failures) on intervention k by time t, the posterior distribution is Beta distribution (by conjugacy)

$$\widetilde{\theta}_k \sim \text{Beta}(s_k(t), f_k(t))$$

where $s_k(t) = s_k(0) + s_k, f_k(t) = f_k(0) + f_k$

DP Procedure

- We use dynamic programming (DP) to obtain an optimal adaptive intervention allocation sequence
- Bayes-optimal means maximising the Bayes-expected total number of successes (patient benefit) in the trial
- Specifically, we use backward recursion algorithm

DP Procedure: Computational Tractability

Publication	T	$T^{\max}$	SW, HW, RAM
Steck (1964)	25	N/A	N/A, UNIVAC 1105, 54 kB
Yakowitz (1969)	5	N/A	Fortran, N/A, N/A
Berry (1978)	100	N/A	Basic (?), Atari (?), N/A
Ginebra and Clayton (1999)	150	180	N/A, N/A, N/A
Hardwick et al. (2006)	100	200	N/A, N/A, N/A
Ahuja and Birge (2016)	96	240	N/A, Mac 4GB
Williamson et al. (2017)	100	215	R, PC, 16GB
Villar (2018)	100	N/A	Matlab, PC, N/A
Kaufmann (2018)	70	N/A	N/A, N/A, N/A

DP Procedure: Computational Tractability

Publication	T	$T^{\max}$	SW, HW, RAM
Steck (1964)	25	N/A	N/A, UNIVAC 1105, 54 kB
Yakowitz (1969)	5	N/A	Fortran, N/A, N/A
Berry (1978)	100	N/A	Basic (?), Atari (?), N/A
Ginebra and Clayton (1999)	150	180	N/A, N/A, N/A
Hardwick et al. (2006)	100	200	N/A, N/A, N/A
Ahuja and Birge (2016)	96	240	N/A, Mac 4GB
Williamson et al. (2017)	100	215	R, PC, 16GB
Villar (2018)	100	N/A	Matlab, PC, N/A
Kaufmann (2018)	70	N/A	N/A, N/A, N/A
Jacko (2019a)	4440	4440	Julia 1.0.1 & BB, PC, 32GB

DP P… …ctability

Publicat…

Steck (1… …105, 54 kB
Yakowitz… …/A
Berry (1… …(?), N/A
Ginebra …
Hardwic…
Ahuja a…
Williams…
Villar (2… …A
Kaufma…
Jacko (2… …B, PC, 32GB

Please cite this paper as:

Jacko, P. (2019). BinaryBandit: An Efficient Julia Package for Optimization and Evaluation of the Finite-Horizon Bandit Problem with Binary Responses. Management Science Working Paper 2019:4, Lancaster University Management School, 13 pages.

Lancaster University
Management School

Management Science

Working Paper 2019:4

BinaryBandit: An Efficient Julia Package for Optimization and Evaluation of the Finite-Horizon Bandit Problem with Binary Responses

Peter Jacko

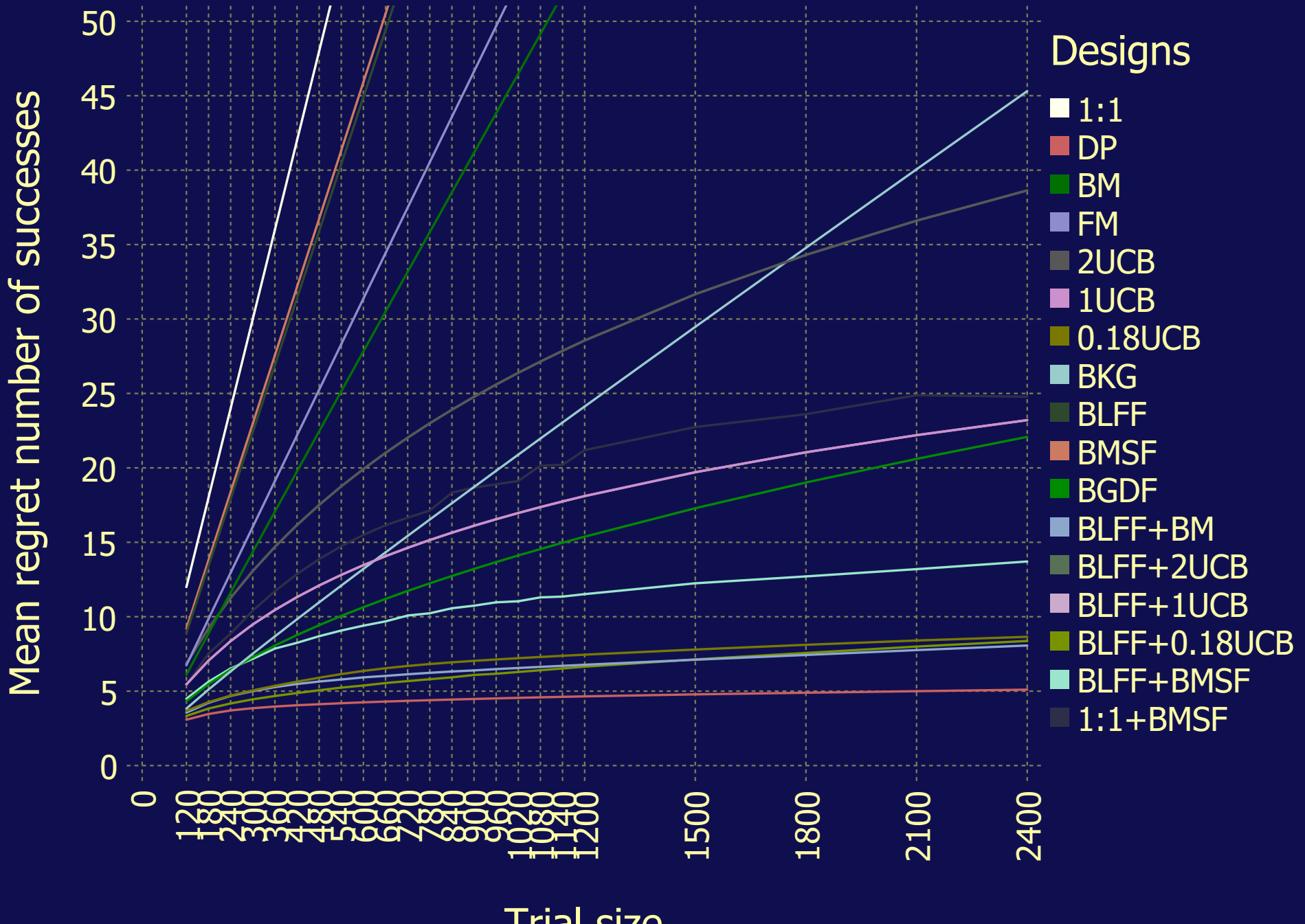
Frequentist UCB Index Rules

- Arm with highest upper confidence bound gets priority
  - ▷ either it has high sample mean
  - ▷ or it has high uncertainty around the mean
- Many variants, computationally notably different
- α-UCB (originally $\alpha = 2$ in Auer et al. (2002)):

$$\frac{s_k(t)}{s_k(t) + f_k(t)} + \sqrt{\frac{\alpha \cdot \ln(t+1)}{s_k(t) + f_k(t)}}$$

  - ▷ currently theoretical performance bounds for $\alpha > 1$
  - ▷ $\alpha = 1$ often used in computational papers
  - ▷ $\alpha = 0.18$ found numerically as best performing

Procedures: Frequentist Regret



Procedures: Frequentist Regret

The Finite-Horizon Two-Armed Bandit Problem with Binary Responses

A Multidisciplinary Survey of the History, State of the Art, and Myths

Peter Jacko
Department of Management Science
Lancaster University, UK

June 18, 2019

Abstract

In this paper we consider the two-armed bandit problem, which often naturally appears per se or as a subproblem in some multi-armed generalizations, and serves as a starting point for introducing additional problem features. The consideration of binary responses is motivated by its widespread applicability and by being one of the most studied settings. We focus on the undiscounted finite-horizon objective, which is the most relevant in many applications. We make an attempt to unify the terminology as this is different across disciplines that have considered this problem, and present a unified model cast in the Markov decision process framework, with subject responses modelled using the Bernoulli distribution, and the corresponding Beta distribution for Bayesian updating. We give an extensive account of the history and state of the art of approaches from several disciplines, including design of experiments, Bayesian decision theory, naive designs, reinforcement learning, biostatistics, and combination designs. We evaluate these designs, together with a few newly proposed, accurately computationally (using a newly written package in Julia programming language by the author) in order to compare their performance. We show that conclusions are different for moderate horizons (typical in practice) than for small horizons (typical in academic literature reporting computational results). We further list and clarify a number of myths about this problem, e.g., we show that, computationally, much larger problems can be designed to Bayes-optimality than what is commonly believed.

Deterministic Procedures

- Problem? Deterministic procedures are not suitable to implement in many clinical trials because randomisation is required to avoid several sources of bias
- Therefore, we need to modify the DP procedure by forcing actions to be randomised

Deterministic Procedures

- Problem? ...e to implement randomisa... of bias
- Therefore, ...y forcing ac...

Biometrika (2007), **94**, 3, *pp.* 673–689
© 2007 Biometrika Trust
Printed in Great Britain

doi:10.1093/biomet/asm049
Advance Access publication 5 August 2007

Optimal adaptive randomized designs for clinical trials

BY YI CHENG

Department of Mathematical Sciences, Indiana University, South Bend, Indiana 46634, U.S.A.

ycheng@iusb.edu

AND DONALD A. BERRY

Biostatistics Department, The University of Texas, M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.

dberry@mdanderson.org

SUMMARY

Optimal decision-analytic designs are deterministic. Such designs are appropriately criticized in the context of clinical trials because they are subject to assignment bias. On the other hand, balanced randomized designs may assign an excessive number of patients to a treatment arm that is performing relatively poorly. We propose a compromise between these two extremes, one that achieves some of the good characteristics of both. We introduce a constrained optimal adaptive design for a fully sequential randomized clinical trial with k arms and n patients. An r-design is one for which, at each allocation, each arm has probability at least r of being chosen, $0 \leqslant r \leqslant 1/k$. An optimal design among all r-designs is called r-optimal. An r_1-design is also an r_2-design if $r_1 \geqslant r_2$. A

Randomised DP

- Action 1: intervention C is allocated with probability p
- Action 2: intervention D is allocated with probability p
- The expected total number of successes under Action 1

$$\mathcal{V}_m^1(\boldsymbol{z}) = p \cdot \mathcal{F}_m^C(\boldsymbol{z}) + (1-p) \cdot \mathcal{F}_m^D(\boldsymbol{z})$$

- The objective function becomes

$$\mathcal{V}_m(\boldsymbol{z}) = \max\left\{\mathcal{V}_m^1(\boldsymbol{z}),\ \mathcal{V}_m^2(\boldsymbol{z})\right\}$$

- Lower biases, but lower controllability

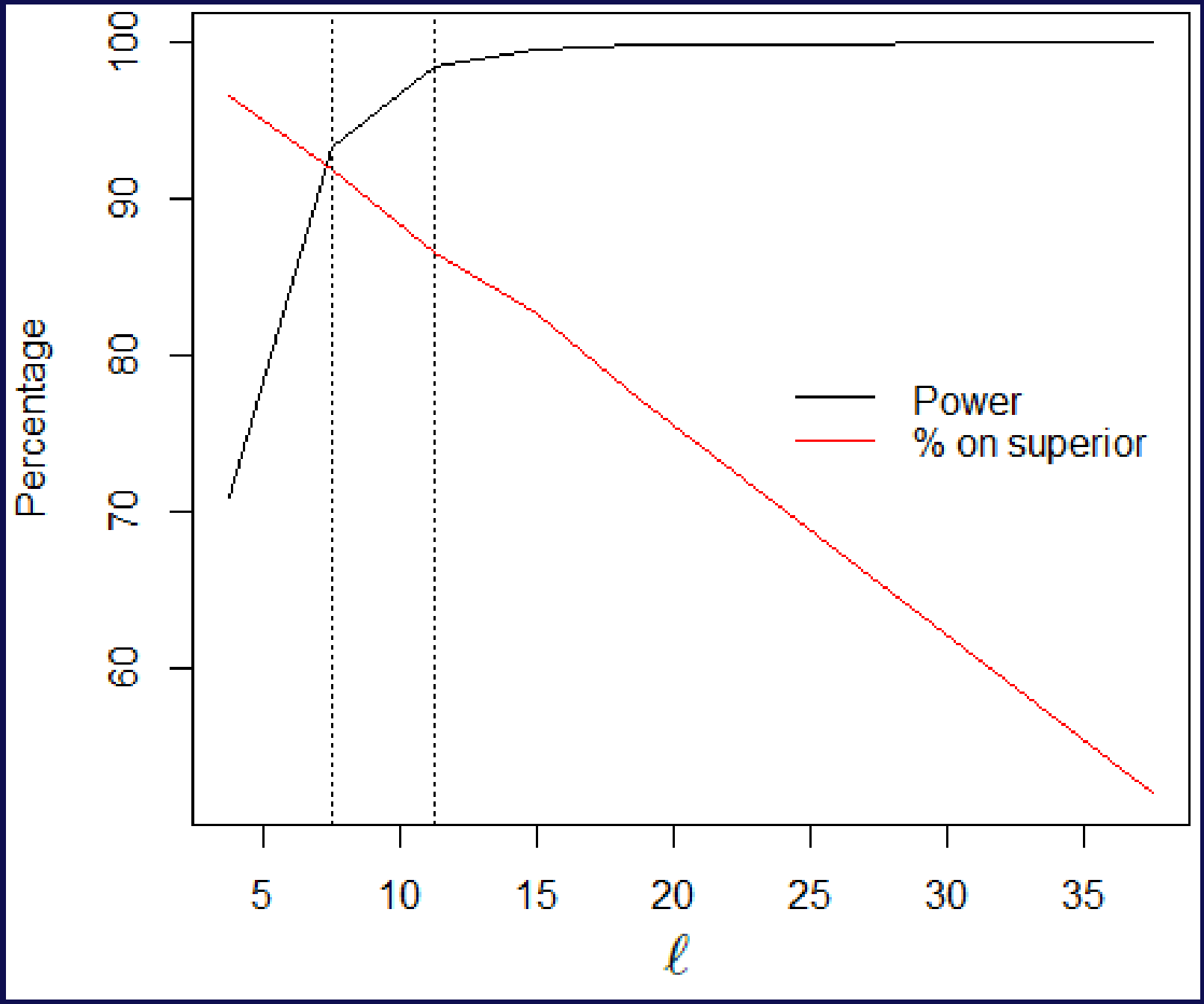
Randomised DP

- Problems? After running simulations, we found:
  - ▷ this procedure is very underpowered for high p
  - ▷ in some of the runs, all subjects were allocated to only one of interventions
- This means we cannot be confident about the results
- ...we cannot calculate important performance measures
- Therefore, we lower-limit the number of observations on each intervention

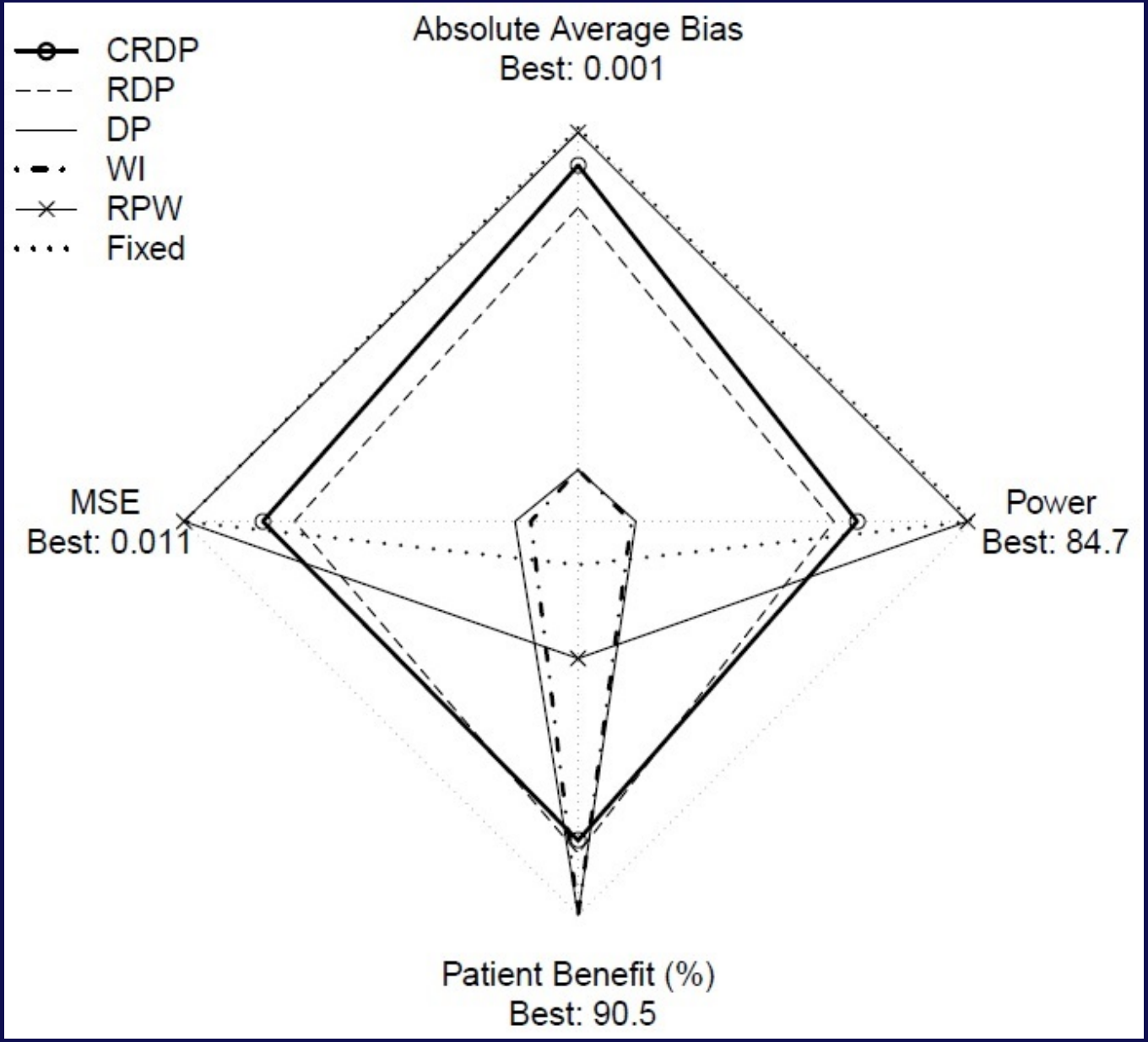
Constrained Randomised DP

- We modify the optimal randomised DP procedure by adding a constraint to ensure that we obtain $\geq \ell$ observations from each intervention
- To do this, we assign a large penalty to every terminal state that has $< \ell$ observations on an arm
- The undesirable states will now be avoided (as much as possible) by the procedure
- We suggested $p = 0.9$ and $\ell = 0.15T$
  - ▷ Note that $0.50T$ corresponds to 1:1

Constrained Randomised DP



Comparison


CRDP
RDP
DP
WI
RPW
Fixed
Absolute Average Bias
Best: 0.001
MSE
Best: 0.011
Power
Best: 84.7
Patient Benefit (%)
Best: 90.5

Comparison

Computational Statistics and Data Analysis 113 (2017) 136–153

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

r
4.7

A Bayesian adaptive design for clinical trials in rare diseases

S. Faye Williamson [a,*], Peter Jacko [b], Sofía S. Villar [c], Thomas Jaki [a]

[a] Department of Mathematics and Statistics, Lancaster University, UK
[b] Department of Management Science, Lancaster University, UK
[c] MRC Biostatistics Unit, Cambridge, UK

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form 9 September 2016
Accepted 10 September 2016
Available online 28 September 2016

Keywords:
Clinical trials
Rare diseases

ABSTRACT

Development of treatments for rare diseases is challenging due to the limited number of patients available for participation. Learning about treatment effectiveness with a view to treat patients in the larger outside population, as in the traditional fixed randomised design, may not be a plausible goal. An alternative goal is to treat the patients within the trial as effectively as possible. Using the framework of finite-horizon Markov decision processes and dynamic programming (DP), a novel randomised response-adaptive design is proposed which maximises the total number of patient successes in the trial and penalises if a minimum number of patients are not recruited to each treatment arm. Several

Computational Statistics and Data Analysis 174 (2022) 107407

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Generalisations of a Bayesian decision-theoretic randomisation procedure and the impact of delayed responses

S. Faye Williamson [a,b,*], Peter Jacko [c,d], Thomas Jaki [a,e]

[a] Department of Mathematics and Statistics, Lancaster University, UK
[b] Biostatistics Research Group, Population Health Sciences Institute, Newcastle University, UK
[c] Department of Management Science, Lancaster University, UK
[d] Berry Consultants, UK
[e] MRC Biostatistics Unit, Cambridge University, UK

ARTICLE INFO

Article history:
Received 29 June 2021
Received in revised form 30 November 2021
Accepted 1 December 2021
Available online 7 December 2021

Keywords:

ABSTRACT

The design of sequential experiments and, in particular, randomised controlled trials involves a trade-off between operational characteristics such as statistical power, estimation bias and patient benefit. The family of randomisation procedures referred to as Constrained Randomised Dynamic Programming (CRDP), which is set in the Bayesian decision-theoretic framework, can be used to balance these competing objectives. A generalisation and novel interpretation of CRDP is proposed to highlight its inherent flexibility to adapt to a variety

A Bayesian ada

S. Faye Williamson

[a] Department of Mathematics a
[b] Department of Management Science, Lancaster University, UK
[c] MRC Biostatistics Unit, Cambridge, UK

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form 9 September 2016
Accepted 10 September 2016
Available online 28 September 2016

Keywords:
Clinical trials
Rare diseases

ABSTRACT

Development of treatments for rare diseases is challenging due to the limited number of patients available for participation. Learning about treatment effectiveness with a view to treat patients in the larger outside population, as in the traditional fixed randomised design, may not be a plausible goal. An alternative goal is to treat the patients within the trial as effectively as possible. Using the framework of finite-horizon Markov decision processes and dynamic programming (DP), a novel randomised response-adaptive design is proposed which maximises the total number of patient successes in the trial and penalises if a minimum number of patients are not recruited to each treatment arm. Several

Comparison $T = 148$: $H_0 : \theta_C = \theta_D = 0.3$

	z-test 0.95/0.98	F-test 0.91/0.95	EPASA (SD)
EFR	0.051/0.021	0.058/0.024	0.500 (0.041)
LFF	0.054/0.023	0.057/0.024	0.500 (0.029)
2UCB	0.063/0.031	0.068/0.033	0.500 (0.101)
0.5UCB	0.089/0.049	0.095/0.050	0.500 (0.199)
0.18UCB	0.091/0.047	0.101/0.047	0.500 (0.308)
0UCB	0.001/0.000	0.001/0.000	0.500 (0.483)
37C+0.8RDP	0.063/0.030	0.068/0.031	0.500 (0.181)
15C+0.95RDP	0.091/0.048	0.101/0.049	0.500 (0.298)
0.99RDP	0.077/0.031	0.097/0.034	0.500 (0.344)
37C+DP	0.063/0.030	0.068/0.031	0.500 (0.209)
15C+DP	0.092/0.047	0.105/0.047	0.500 (0.313)
7C+DP	0.089/0.029	0.116/0.032	0.500 (0.343)
DP	0.073/0.026	0.094/0.028	0.500 (0.352)
WI	0.065/0.022	0.090/0.024	0.500 (0.363)
Oracle	0.000/0.000	0.000/0.000	0.500 (0.500)

Comparison $T = 148$:

$H_1 : \theta_C = 0.3\ ,\ \theta_D = 0.5$

	z-test 00.95/0.98	F-test 0.91/0.95	EPASA (SD)	ENS (SD)
EFR	0.805/0.676	0.755/0.589	0.500 (0.041)	59.200 (5.960)
LFF	0.804/0.672	0.746/0.567	0.586 (0.033)	61.735 (6.199)
2UCB	0.786/0.637	0.707/0.497	0.727 (0.077)	65.915 (6.543)
0.5UCB	0.650/0.442	0.547/0.308	0.838 (0.103)	69.219 (6.894)
0.18UCB	0.356/0.158	0.308/0.104	0.877 (0.163)	70.356 (7.740)
0UCB	0.012/0.007	0.011/0.004	0.692 (0.445)	64.883 (14.51)
37C+0.8RDP	0.746/0.600	0.663/0.478	0.714 (0.060)	65.527 (6.240)
15C+0.95RDP	0.580/0.412	0.504/0.314	0.840 (0.118)	69.270 (7.021)
0.99RDP	0.323/0.170	0.308/0.123	0.882 (0.166)	70.504 (7.849)
37C+DP	0.715/0.575	0.634/0.461	0.734 (0.050)	66.128 (6.159)
15C+DP	0.536/0.376	0.467/0.288	0.854 (0.114)	69.666 (6.962)
7C+DP	0.411/0.250	0.369/0.219	0.880 (0.151)	70.441 (7.590)
DP	0.263/0.116	0.262/0.078	0.888 (0.172)	70.696 (7.964)
WI	0.233/0.102	0.240/0.069	0.887 (0.184)	70.667 (8.185)
Oracle	0.000/0.000	0.000/0.000	1.000 (0.000)	74.000 (6.083)

Conclusion

- Many important contributions from a number of disciplines over 90 years, but still not clear what the "best" procedure is
  - ▷ and this is just the simplest problem variant!
- Many myths (see Jacko (2019b))
- Randomised procedures in biostatistics — a whole new world
- Unclear computational limits of DP and Gittins/Whittle index

Thank you for your attention

Ďakujem za Vašu pozornosť

Links

- Jacko (2019b): https://arxiv.org/pdf/1906.10173.pdf
- Jacko (2019a): https://eprints.lancs.ac.uk/id/eprint/136340/1/Jacko2019_binarybandit_wp.pdf
- BinaryBandit Julia package: https://github.com/PeterJacko/BinaryBandit
- R ShinyApp: https://peterjacko.shinyapps.io/binarybandit-app/
- Group on Optimal Adaptive Learning (G.O.A.L.): http://www.lancaster.ac.uk/staff/jacko/goal/

References

Ahuja, V. and Birge, J. R. (2016). Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *European Journal of Operational Research*, 248:619–633.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256.

Berry, D. A. (1978). Modified two-armed bandit strategies for certain clinical trials. *Journal of the American Statistical Association*, 73:339–345.

Ginebra, J. and Clayton, M. K. (1999). Small-sample performance of Bernoulli two-armed bandit Bayesian strategies. *Journal of Statistical Planning and Inference*, 79(1):107–122.

Hardwick, J., Oehmke, R., and Stout, Q. F. (2006). New adaptive designs for delayed response models. *Journal of Statistical Planning and Inference*, 136:1940–1955.

Jacko, P. (2019a). BinaryBandit: An efficient Julia package for optimization and evaluation of the finite-horizon bandit problem with binary responses. Management Science Working Paper 2019:4, Lancaster University Management School.

Jacko, P. (2019b). The finite-horizon two-armed bandit problem with binary responses: A multidisciplinary survey of the history, state of the art, and myths. Management Science Working Paper 2019:3, Lancaster University Management School. arXiv:1906.10173.

Kaufmann, E. (2018). On Bayesian index policies for sequential resource allocation. *The Annals of Statistics*, 46(2):842–865.

Steck, R. (1964). A dynamic programming strategy for the two machine problem. *Mathematics of Computation*, 18(86):285–291.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

Villar, S. S. (2018). Bandit strategies evaluated in the context of clinical trials in rare life-threatening diseases. *Probability in the Engineering and Informational Sciences*, 32:229–245.

Villar, S. S. and Jacko, P. (2022). Bandit procedures for designing patient-centric clinical trials. In *The Elements of Joint Learning and Optimization in Operations Management*, pages 365–389. Springer. Invited book chapter.

Williamson, S. F., Jacko, P., and Jaki, T. (2022). Generalisations of a Bayesian decision-theoretic randomisation procedure and the

impact of delayed responses. *Computational Statistics and Data Analysis*, 174:107407.

Williamson, S. F., Jacko, P., Villar, S. S., and Jaki, T. (2017). A Bayesian adaptive design for clinical trials in rare diseases. *Computational Statistics and Data Analysis*, 113C:136–153.

Yakowitz, S. J. (1969). *Mathematics of Adaptive Control Processes*. New York, NY: North-Holland.