The Efficient Covariate-Adaptive Design for high-order balancing of quantitative and qualitative covariates

The Efficient Covariate-Adaptive Design for high-order balancing of quantitative and qualitative covariates

Alessandro Baldi Antognini[1*], Rosamarie Frieri[1], Maroussa Zagoraiou[1] and Marco Novelli[1]

[1]Department of Statistical Sciences, University of Bologna, via delle Belle Arti 41, Bologna, 40126, Italy.

*Corresponding author(s). E-mail(s): a.baldi@unibo.it;
Contributing authors: rosamarie.frieri2@unibo.it;
maroussa.zagoraiou@unibo.it; m.novelli@unibo.it;

Abstract

In the context of sequential treatment comparisons, the acquisition of covariate information about the statistical units is crucial for the validity of the trial. Furthermore, balancing the assignments among covariates is of primary importance, since the potential imbalance of the covariate distributions across the groups can severely undermine the statistical analysis. For this reason, several covariate-adaptive randomization procedures have been suggested in the literature, but most of them only apply to categorical factors. In this paper we propose a new class of rules, called the Efficient Covariate-Adaptive Design, which is high-order balanced regardless of the number of factors and their nature (qualitative and/or quantitative), also accounting for every order covariate effects and interactions. The suggested procedure performs very well, is flexible and simple to implement. The advantages of our proposal are also analyzed via simulations and its finite sample properties are compared with those of other well-known rules, by also including the redesign of a real clinical trial.

Keywords: Biased Coin Design, Markov chains, Minimization Methods, Stratified Randomization.

1 Introduction

From Fisher around 1930, randomization has been the fundamental tool to evaluate the effects of two treatments, say A and B, that must be compared. Randomization is now the gold standard for clinical experiments and its popularity is growing in every applied field, since a random allocation to the different arms is recommended to build a solid base for inference and make the treatment groups comparable (at least asymptotically) with respect to unknown covariates, also mitigating the accidental bias due to unknown confounders and the selection bias induced by the investigators. At the same time, from the inferential viewpoint balance is *universally optimal* under the linear homoscedastic model assumptions (see for a review Baldi Antognini and Giovagnoli (2015)). To obtain a valid compromise between balance and randomness, Efron (1971) introduced his well-known Biased Coin Design (BCD), namely a sequential allocation rule favoring at each step the under-represented treatment. By selecting the difference D_n between the assignments to A and B as the imbalance measure after n allocations, Efron's rule randomizes the $(n+1)$th patient to A with probability $h_E(D_n)$, where

$$h_E(x) = 1/2 - \text{sgn}(x)\,(\rho - 1/2) \;\text{ with }\; \rho \in (1/2;\ 1).$$

The peculiarities of this proposal are that $h_E(\cdot)$ is decreasing with $h_E(x) = h_E(\theta x)$ for any $\theta > 0$. Under Efron's coin, $\{D_n\}_{n\in\mathbb{N}}$ is an irreducible and positive recurrent Markov chain and so it is bounded in probability, namely $D_n = O_p(1)$. This guarantees the highest order of convergence with respect to the other BCDs proposed by Wei (1978), Atkinson (1982) and Smith (1984a,b), that are based on continuous randomization functions - in place of $h_E(\cdot)$ - applied to imbalance measures re-scaled by n (like, e.g., $n^{-1}D_n$), under which $D_n = O_p(\sqrt{n})$.

However, in the large majority of comparative trials, patients' heterogeneity is taken into account by including covariates/prognostic factors in the design phase as well as in the analysis and one of the fundamental goals consists in generating comparable groups with respect to the chosen set of covariates. Although sometimes vague in practice, the concept of covariate balance should provide treatment arms as much as possible homogeneous in terms of the covariate distribution. This enhances the credibility of the analysis, making it more robust against model misspecification, since a significant degree of covariate imbalance could severely undermine the inferential efficiency about the treatment effects. This demand is particularly cogent for sequential experiments, where keeping a reasonable degree of balance at any time could allow to stop the experiment under an excellent inferential setting. Even if balance is commonly considered desirable from several viewpoints, its mathematical justification is strictly related to the linear homoscedastic model. Indeed, suppose that for each statistical unit we observe a p-dim vector $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^t$ of covariates assumed to be random but measurable before the treatment assignments. Suppose that n allocations of either treatment A or B have been made to patients with independent and identically distributed covariates $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \sim \mathcal{L}(\boldsymbol{z})$, where $\mathcal{L}$ denotes their common p-variate (joint) probability distribution/density function. Moreover, assume that the outcomes $\boldsymbol{Y}_n = (Y_1, \ldots, Y_n)^t$ follow (at least approximately) the

linear model:

$$\mathbb{E}(\boldsymbol{Y}_n) = \boldsymbol{\delta}_n \mu_A + (\mathbf{1}_n - \boldsymbol{\delta}_n)\, \mu_B + \mathbb{F}_n \boldsymbol{\beta}, \quad \text{var}(\boldsymbol{Y}_n) = \sigma^2 \mathbb{I}_n, \tag{1}$$

where $\mathbb{I}_n$ and $\mathbf{1}_n$ denote the $n-$dim identity matrix and the n-dim (column) vector of ones, $\boldsymbol{\delta}_n = (\delta_1, \ldots, \delta_n)^t$ is the allocation, with $\delta_i = 1$ if the ith subject is assigned to A and 0 otherwise, $\boldsymbol{\beta}$ is a q-dim vector of covariate effects, $\mathbb{F}_n = [\mathbf{f}(\boldsymbol{Z})^t]$ with $\mathbf{f}(\cdot)$ a known q-dim vector function (usually $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$, but it may also include higher order terms or interactions among the covariates so that, in general, $q \geq p$) and σ^2 is the common variance. Since model (1) does not account for treatment-covariate interaction, it is customary to regard $\boldsymbol{\beta}$ as a nuisance parameter and the inferential goal lies in estimating $\boldsymbol{\mu} = (\mu_A;\ \mu_B)^t$ or $\mu_A - \mu_B$ as precisely as possible. Let $\bar{\mathbf{f}}_n = n^{-1}\mathbb{F}_n^t \mathbf{1}_n$ and $\mathbb{A}_n = n^{-1}\mathbb{F}_n^t \mathbb{F}_n$, if $\mathbb{P}_n = \begin{pmatrix} 1 & \bar{\mathbf{f}}_n^t \\ \bar{\mathbf{f}}_n & \mathbb{A}_n \end{pmatrix}$ is non-singular, from the optimal design theory the design efficiency for estimating $\boldsymbol{\mu}$ with respect to both A and D-optimality is $1 - n^{-1}\ell_n$, where $\ell_n = n^{-1}\mathbf{b}_n^t \mathbb{P}_n^{-1} \mathbf{b}_n$ represents the loss of estimation precision induced by the covariate imbalance after n assignments, while $\mathbf{b}_n^t = (D_n;\ (2\boldsymbol{\delta}_n - \mathbf{1}_n)^t \mathbb{F}_n)$ is the so-called imbalance vector. Thus, the estimation efficiency depends on the design only through the loss and it is maximized when $\ell_n = 0$. The same conclusion holds when the inferential interest lies in estimating/testing the treatment difference $\mu_A - \mu_B$; indeed, the design efficiency is still $1 - n^{-1}\ell_n$ and the power of the Wald statistic for testing $H_0 : \mu_A = \mu_B$ vs $H_1 : \mu_A \neq \mu_B$ is maximized when $\ell_n = 0$ (see Appendix A.1).

Several allocation rules have been suggested for balancing the treatments among the set of covariates of interest, while preserving a randomization component in the assignments. Following Rosenberger and Lachin (2002), these procedures fall into the class of Covariate-Adaptive (CA) randomization, since the allocation probability at each step depends on the assignments and the covariates of previous subjects and on the current patient's profile. Inspired by Wynn's algorithm for the sequential construction of D-optimal designs, Atkinson (1982) introduced his D_A-BCD by assuming that the probability of assigning A to the $(n+1)$th unit with profile $\boldsymbol{z}_{n+1}$ is a continuous and decreasing function of the imbalance measure $n^{-1}\,(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\,\mathbb{P}_n^{-1}\mathbf{b}_n$ (see Begg and Iglewicz (1980) and Smith (1984a,b)). Even if it accounts for any type of covariates, the D_A-BCD is quite complex and theoretical results are available only for categorical factors, where the stratum imbalance process is of order $O_p(\sqrt{n})$ (Baldi Antognini and Zagoraiou, 2017).

In the presence of solely categorical covariates, stratified randomization and marginal procedures were introduced to achieve stratum and marginal balance, respectively. The former generates a separate randomization sequence within each stratum, so each subject is randomized according to the evolution of the allocations of his/her profile. If employed via Efron's coin, the stratum imbalance process is still bounded in probability (Baldi Antognini and Zagoraiou, 2011), guaranteeing a higher order of convergence with respect to the D_A-BCD. However, stratified methods are less efficient in the presence of a large number of factors/levels; in such a case, marginal procedures are particularly effective, since they are intended to achieve the less restrictive requirement of marginal balance (Taves, 1974; Pocock and Simon,

1975; Wei, 1978; Heritier et al, 2005). When a subject is ready to be randomized, a weighted sum of all marginal imbalances corresponding to his/her profile is computed, then the allocation is forced to the under-represented treatment (for instance, by using Efron's coin as in Pocock and Simon (1975)). By combining marginal procedures and stratified randomization, Hu and Hu (2012) proposed a family of CA rules under which the allocations are randomized according to $h_E(\cdot)$ applied to an overall imbalance measure that weighs the roles of the global, marginal and stratum imbalances. Under restrictive and hard to check conditions on the weights, the stratum imbalance process is shown to be bounded in probability.

A critical aspect of stratified randomization, marginal procedures and Hu and Hu's rule is that they can only be applied to categorical covariates, so quantitative factors must be discretized. Besides the potential bias induced by the subjective choice of the thresholds, discretization generally induces a consistent loss of precision, especially when the covariate distribution is unknown (Atkinson, 2002). As also stated by Ciolino et al (2011) "attempts should be made to balance known prognostic continuous covariates at the design phase". The literature about CA rules for continuous covariates is quite limited, even more so for the mixed scenario of qualitative and quantitative factors, and most of the existing works assess the performance of the proposed procedures only by simulations. One of the few exceptions is the rerandomization (RR) suggested by Morgan and Rubin (2012) for continuous or binary covariates: by assuming the Mahalanobis distance as the imbalance measure, RR repeatedly randomizes the units via complete randomization until the imbalance is lower than a prefixed threshold. Although introduced in the context of causal inference for non-sequential experiments (with all units immediately available before randomization), RR has been later generalized in a group sequential fashion (Zhou et al, 2018).

The aim of this paper is to introduce a new class of CA randomization rules, the Efficient Covariate-Adaptive DEsign (ECADE), to cope with imbalances in the allocations of two competing treatments that i) can be applied to qualitative and/or quantitative factors, ii) is high-order balanced, that ensures excellent performance in terms of inferential precision and iii) is simple to implement and performs very well even in the presence of a large number of covariates. In particular, the ECADE is based on the sequential minimization of a suitably chosen weighted Euclidean norm of the imbalance vector. We theoretically prove that the underlined imbalance process is a Markov chain bounded in probability and preserves the order of $O_p(1)$, regardless of the nature of the chosen covariates. This guarantees that the loss of estimation efficiency as well as the Mahalanobis distance (which are asymptotically equivalent) tend to zero asymptotically, independently of the number of considered factors. According to the choice of the weights and the allocation function, interesting cases of the ECADE are investigated and already existing designs are retrieved as special cases, which also allows to simplify some strict conditions in Hu and Hu (2012). Furthermore, we perform a simulation study and we redesign a real clinical trial to stress the validity of our theoretical results and to compare the finite sample properties of the ECADE with those of other well-known CA procedures. Starting

from some preliminaries in Section 2, Section 3 introduces the ECADE and its theoretical properties. Section 4 deals with the finite sample comparisons between CA rules taking also into account the group sequential RR and Section 5 reports the redesign of a real study. Section 6 discusses some practical implications while all the proofs can be found in the Appendix.

Throughout the paper, $\mathbb{J}_n$ is the $n-$dim exchange matrix (with 1 on the anti-diagonal and 0 elsewhere), while $\mathbf{0}_n$ is the n-dim vector of zeros. Given a n-dim vector $\boldsymbol{x}$ and a n-dim matrix $\mathbb{W}$ symmetric and positive-definite, $\|\boldsymbol{x}\|_{\mathbb{W}} = \sqrt{\boldsymbol{x}^t\mathbb{W}\boldsymbol{x}}$ is the weighted Euclidean norm of $\boldsymbol{x}$ with respect to $\mathbb{W}$; if $\mathbb{W} = \mathbb{I}_n$ we simply let $\|\boldsymbol{x}\|$ be the Euclidean norm, while $\|\boldsymbol{x}\|_\infty = \max_{i=1,\dots,n}|x_i|$ is the infinity norm. Moreover, let E be an event and E^c its complement, $I\{E\}$ is the indicator function of E, $\otimes$ is the Kronecker product, while $\mathbf{e}_1, \mathbf{e}_2, \dots$ are the canonical basis.

2 Covariate balance and imbalance measures

Even if the concept of covariate balance is sometimes vague, from an inferential viewpoint its meaning, as well as its implications, are linked on the model assumptions. Let $n_A = \sum_{i=1}^n \delta_i$ and $n_B = n - n_A$ be the assignments to A and B after n steps, then $\mathbf{b}_n^t = (D_n;\ (n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B)^t)$, where $\bar{\mathbf{f}}_n^A$ and $\bar{\mathbf{f}}_n^B$ are the vectors of the sample means of $\mathbf{f}(\boldsymbol{Z})$ in the two groups, and

$$\ell_n = n^{-1}\|\mathbf{b}_n\|^2_{\mathbb{P}_n^{-1}} = n^{-1}\|n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}} + \frac{\left[D_n - (n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B)^t\mathbb{A}_n^{-1}\bar{\mathbf{f}}_n\right]^2}{n\left(1 - \|\bar{\mathbf{f}}_n\|^2_{\mathbb{A}_n^{-1}}\right)}. \tag{2}$$

Ideally, $\ell_n = 0$ for any sample size n, but this requirement is extremely stringent, especially when p is large, since it involves the finite sample distributions of the covariates through the random entry of the subjects, as well as the random allocations of the treatments. Indeed, besides the global balance D_n, the form of the imbalance vector depends on both the nature of the covariates and the model assumptions via $\mathbf{f}(\cdot)$. For quantitative factors, if $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$ then $n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B$ simply denotes the vector of the difference between the covariate sums in the two groups; while in the presence of interactions, it also accounts for the difference between the cross-products in the treatment arms. As regards qualitative factors, when $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$, $n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B$ is the vector of the marginal imbalances of the covariates at each of the non-reference categories; while in the presence of interactions $n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B$ also accounts for the stratum imbalances. The mixed case, namely when some covariates are qualitative and other quantitative, could be naturally encompassed by combining the two previously discussed scenarios.

A design which is covariate-balanced, namely such that

$$\mathbf{b}_n = \mathbf{0}_{q+1} \quad \Leftrightarrow \quad D_n = 0 \quad \text{and} \quad \bar{\mathbf{f}}_n^A = \bar{\mathbf{f}}_n^B, \tag{3}$$

is optimal for model (1) and it is still optimal for the linear model that accounts for treatment-covariate interactions (Baldi Antognini and Zagoraiou, 2011, 2012). For quantitative factors, when $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$ condition (3) corresponds to the equality of

the sample covariate means in the two groups and, if we also consider first-order interactions, it means equal covariances in the treatment arms. Condition (3) leads to the equality of all the empirical moments as the model complexity grows. For categorical factors, with no interactions among covariates condition (3) states that A and B are equally replicated at every level of each blocking factor (the marginal balance $\mathcal{L}_n^A(Z_k) = \mathcal{L}_n^B(Z_k)$ for $k = 1, \ldots, p$); while if the model is full (namely with main effects and interactions of all orders), A and B should be equally replicated also within every stratum, leading to $\mathcal{L}_n^A(\boldsymbol{Z}) = \mathcal{L}_n^B(\boldsymbol{Z})$.

Remark 1 Assuming $\mathbb{E}[\mathbf{f}(\boldsymbol{Z})] = \boldsymbol{\mu}_{\mathbf{f}} < \infty$ and $\mathbb{E}\left[\mathbf{f}(\boldsymbol{Z})\mathbf{f}(\boldsymbol{Z})^t\right] = \lim_{n\to\infty} \mathbb{A}_n = \mathbb{A}$ non-singular, from the strong law of large numbers

$$\lim_{n\to\infty} \mathbb{P}_n = \mathbb{P} = \begin{pmatrix} 1 & \boldsymbol{\mu}_{\mathbf{f}}^t \\ \boldsymbol{\mu}_{\mathbf{f}} & \mathbb{A} \end{pmatrix} \quad a.s. \tag{4}$$

regardless of the chosen design. When $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$, $\mathbb{A}_n$ is the sample second moment matrix of the covariates; thus, (4) holds with $\mathbb{A} = \mathbb{V} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^t$, where $\mathbb{V} = \text{var}(\boldsymbol{Z})$.

Even if the loss plays a primary role, other scalar measures of overall covariate imbalance have been proposed in the literature. These indicators usually combine the observed imbalances in all the considered factors by taking into account their specific relevance in the trial, the nature of the covariates (categorical/quantitative) and, eventually, their dependence structure. For instance, in the case of qualitative covariates Hu and Hu (2012) assumed a weighted mean of the global imbalance, the marginal imbalances and the stratum one. Whereas in the context of RR for causal inference, Morgan and Rubin (2012) chose the Mahalanobis distance between the covariate means in the two groups as the overall measure of covariate imbalance, namely $\mathcal{M}_n = (\bar{\mathbf{f}}_n^A - \bar{\mathbf{f}}_n^B)^t \text{var}(\bar{\mathbf{f}}_n^A - \bar{\mathbf{f}}_n^B)^{-1}(\bar{\mathbf{f}}_n^A - \bar{\mathbf{f}}_n^B) = n^{-1} n_A n_B \|\bar{\mathbf{f}}_n^A - \bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}}$.

Remark 2 As in the original framework of Morgan and Rubin (2012), by assuming that all units and their covariates are immediately available before randomization and setting in advance $n_A = n_B$, if $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$ then $\mathcal{M}_n$ corresponds to a simplified version of the loss. Indeed, in a finite set up with a prefixed n, by centering the covariates we obtain $\bar{\mathbf{f}}_n = \mathbf{0}_q$ so that, from (2), $\ell_n = n^{-1}[D_n^2 + \|n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}}]$ and thus $\ell_n = \mathcal{M}_n$ when $D_n = 0$. The same holds asymptotically, where ℓ_n is equivalent to $\mathcal{M}_n$ provided that D_n vanishes as n grows; indeed, for centered covariates $\boldsymbol{\mu}_{\mathbf{f}} = \mathbf{0}_q$ and, from (4), $\|\mathbf{b}_n\|^2_{\mathbb{P}_n^{-1}}$ could be approximated by $D_n^2 + \|n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B\|^2_{\mathbb{V}^{-1}}$ (if $\mathbb{V}$ is singular, $\mathbb{V}^{-1}$ can be replaced by its pseudo-inverse), so that $\ell_n \approx n4^{-1}\|\bar{\mathbf{f}}_n^A - \bar{\mathbf{f}}_n^B\|^2_{\mathbb{V}^{-1}}$. Moreover, ℓ_n and $\mathcal{M}_n$ could be asymptotically equivalent also in a sequential framework (see Theorem 1).

3 The Efficient Covariate-Adaptive Design

If a design is covariate-balanced then every measure of imbalance vanishes. Whereas, for a properly chosen imbalance measure it is possible to implement a sequential strategy that is nearly balanced for every sample size, guaranteeing a high order of convergence. In particular, we introduce a new class of CA rules, called

the Efficient Covariate-Adaptive Design, based on the sequential minimization of $\|\mathbf{b}_n\|^2_{\mathbb{W}}$, where $\mathbb{W}$ is a symmetric and positive-definite weight matrix. Let $\Im_n$ be the sigma-algebra generated by the first n assignments $\boldsymbol{\delta}_n$ and covariates $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ (with $\Im_0$ the trivial sigma-field). When the $(n+1)$th subject with profile $\boldsymbol{z}_{n+1}$ is ready to be randomized, we calculate the potential imbalances $\|\mathbf{b}^{(A)}_{n+1}\|^2_{\mathbb{W}}$ and $\|\mathbf{b}^{(B)}_{n+1}\|^2_{\mathbb{W}}$ corresponding to an allocation to A or B, respectively. Since $\mathbf{b}^{(A)}_{n+1} = \mathbf{b}_n + (1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)^t$ and $\mathbf{b}^{(B)}_{n+1} = \mathbf{b}_n - (1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)^t$, then $\|\mathbf{b}^{(A)}_{n+1}\|^2_{\mathbb{W}} - \|\mathbf{b}^{(B)}_{n+1}\|^2_{\mathbb{W}} = 4(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{W}\mathbf{b}_n$ and the ECADE randomizes this subject according to

$$\Pr(\delta_{n+1} = 1 \mid \Im_n, \boldsymbol{Z}_{n+1} = \boldsymbol{z}_{n+1}) = h\left((1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{W}\mathbf{b}_n\right), \tag{5}$$

where $h : \mathbb{R} \longrightarrow (0,1)$ is a decreasing and symmetric function, with $h(x) = 1 - h(-x)$, and it is strictly decreasing at 0.

Remark 3 Since $\mathbb{W}$ is symmetric and positive-definite, then $\|\mathbf{b}_n\|^2_{\mathbb{W}}$ is a vector norm and therefore $\|\mathbf{b}_n\|^2_{\mathbb{W}} = 0$ if and only if $\mathbf{b}_n = \mathbf{0}_{q+1}$ (while the less stringent requirement of $\mathbb{W}$ positive-semidefinite induces only a semi-norm, so $\|\mathbf{b}_n\|^2_{\mathbb{W}}$ could vanish also when $\mathbf{b}_n \neq \mathbf{0}_{q+1}$).

The ECADE is extremely flexible. Besides extending Efron's allocation function, $h(\cdot)$ can be appropriately set to meet the desired degree of randomness of the specific clinical study. For instance, $h(x)$ could be modeled as a function of the cdf of a normal distribution as we will discuss in Section 5. Moreover, $\mathbb{W}$ can be chosen according to specific demands on the desired imbalance measure (related for instance to the relevance of the considered factors and their dependence structure). Clearly, the adopted weighted matrix naturally induces a corresponding overall imbalance. For instance, by letting $\mathbb{W} = \mathbb{I}_{q+1}$, the ECADE sequentially minimizes the Euclidean norm of $\mathbf{b}_n$, which essentially corresponds to the proposal of Begg and Iglewicz (1980) for binary factors. When the covariate distribution is a-priori known, by setting $\mathbb{W} = \mathbb{P}^{-1}$ the imbalance measure coincides with the one in Smith (1984a). Moreover, the ECADE can also accommodate the case in which the weights could change step by step (namely, assuming $\mathbb{W} := \mathbb{W}_n$), as the following Example shows.

Example 1 By letting $\mathbb{W} = \mathbb{P}^{-1}_{n+1}$, the ensuing imbalance measure is proportional to the difference among the potential losses induced by assigning the $(n+1)$th subject to A and B, respectively. Indeed, $\ell^{(A)}_{n+1} - \ell^{(B)}_{n+1} = 4(n+1)^{-1}(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{P}^{-1}_{n+1}\mathbf{b}_n$ since, from (2),

$$\ell^{(A)}_{n+1} = (n+1)^{-1}\left\{\mathbf{b}_n^t\mathbb{P}^{-1}_{n+1}\mathbf{b}_n + 2(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{P}^{-1}_{n+1}\mathbf{b}_n + (1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{P}^{-1}_{n+1}(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)^t\right\},$$

$$\ell^{(B)}_{n+1} = (n+1)^{-1}\left\{\mathbf{b}_n^t\mathbb{P}^{-1}_{n+1}\mathbf{b}_n - 2(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{P}^{-1}_{n+1}\mathbf{b}_n + (1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{P}^{-1}_{n+1}(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)^t\right\}.$$

Although apparently similar to Atkinson's rule (for which $\mathbb{W} = \mathbb{P}^{-1}_n$), the ECADE is structurally different since i) the D_A-BCD does not employ the knowledge of the covariate profile $\boldsymbol{z}_{n+1}$ of the $(n+1)$th unit to update $\mathbb{P}_n$ and ii) the ECADE randomizes the assignments based on $(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)\mathbb{P}^{-1}_{n+1}\mathbf{b}_n$, thus the scaling constant n^{-1} in Atkinson's imbalance measure

has no role. For these reasons, the ECADE guarantees a faster convergence to balance with respect to the D_A-BCD, as we will show theoretically and via simulations.

We now explore the probabilistic properties underlining the ECADE, showing in particular that $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is a Markov process bounded in probability with $\mathbf{b}_n = O_p(1)$, which implies that both ℓ_n and $\mathcal{M}_n$ vanish asymptotically. To derive the theoretical properties of our proposal, the cases of qualitative and quantitative covariates are treated separately. The extension to the mixed scenario is straightforward.

3.1 Properties of the ECADE for qualitative covariates

To derive the theoretical properties of the ECADE in the case of qualitative factors, we consider a vectorized notation for the strata. In particular, assuming s_k levels for the kth covariate, the combination of each level of every covariate generates $s = \prod_{k=1}^{p} s_k$ different strata (denoted by $j = 1, \ldots, s$). At step i, let $\mathbf{x}_i = (x_{i1}, \ldots, x_{is})^t$ denote the stratum-indicator vector, where $\mathbf{x}_i = \mathbf{e}_j$ if and only if the ith unit belongs to stratum j. Let $\mathbf{p} = (p_1, \ldots, p_s)^t$ be the probability distribution over the strata, where we assume $p_j > 0$ for every $j = 1, \ldots, s$ (otherwise a given stratum could be excluded) and set $\mathbf{P} = \text{diag}(\mathbf{p})$. After n steps, $N_{nj} = \sum_{i=1}^{n} x_{ij}$ and $\hat{p}_{nj} = n^{-1}N_{nj}$ denote the number and the percentage of subjects within the jth stratum, respectively, while $D_{nj} = 2\sum_{i=1}^{n} \delta_i x_{ij} - N_{nj}$ is the corresponding imbalance; also, let $\mathbf{D}_n = (D_{n1}, \ldots, D_{ns})^t$, $\hat{\mathbf{p}}_n = (\hat{p}_{n1}, \ldots, \hat{p}_{ns})^t$ and $\hat{\mathbf{P}}_n = \text{diag}(\hat{\mathbf{p}}_n)$. For any specification of $\mathbf{f}(\boldsymbol{Z})$, there exists a unique $\{0;\ 1\}$-matrix $\tilde{\mathbb{H}}$ such that $E[\mathbf{f}(\boldsymbol{Z})] = \tilde{\mathbb{H}}\mathbf{p}$; let $\mathbb{H}^t = (\mathbf{1}_s;\ \tilde{\mathbb{H}}^t)$, then $\mathbb{H}$ is a $(q+1) \times s$ matrix with full (row) rank, so $\mathbf{b}_n = \mathbb{H}\mathbf{D}_n$, $(1;\ \mathbf{f}(\boldsymbol{z}_{n+1})^t)^t = \mathbb{H}\mathbf{x}_{n+1}$ and $\mathbb{P}_n = \mathbb{H}\hat{\mathbf{P}}_n\mathbb{H}^t$.

Example 2 Let $\boldsymbol{Z} = (Z_1, Z_2)$ binary with levels $\{z_{i0}, z_{i1}\}$ for Z_i $(i = 1, 2)$, generating $s = 4$ strata. By considering (z_{10}, z_{20}) as the first stratum, (z_{10}, z_{21}) as the second one, (z_{11}, z_{20}), as the third one and (z_{11}, z_{21}) as the fourth stratum, then $\boldsymbol{D}_n = (D_{n1}, D_{n2}, D_{n3}, D_{n4})^t$ and
i) if $\mathbf{f}(\mathbf{Z}) = \mathbf{Z}$ (namely $q = 2$), then $\mathbf{b}_n^t = (D_n, D_n(z_{11}), D_n(z_{21}))$ and $\tilde{\mathbb{H}} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$,
ii) if $\mathbf{f}(\boldsymbol{Z}) = (Z_1, Z_2, Z_1 \cdot Z_2)^t$, then $\mathbf{b}_n^t = (D_n, D_n(z_{11}), D_n(z_{21}), D_n(z_{11}, z_{21}))$ and

$$\tilde{\mathbb{H}} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where $D_n(z_{11}) = \sum_{i=1}^{n}(2\delta_i - 1)I\{Z_{1i} = z_{11}\}$ and $D_n(z_{21}) = \sum_{i=1}^{n}(2\delta_i - 1)I\{Z_{2i} = z_{21}\}$ denote the marginal imbalances at the non-reference categories z_{11} and z_{21}, while $D_n(z_{11}, z_{22}) = D_{n4} = \sum_{i=1}^{n}(2\delta_i - 1)I\{Z_{1i} = z_{11}, Z_{2i} = z_{21}\}$ is the corresponding stratum imbalance.

When the $(n+1)$th unit falling into the jth stratum is ready to be randomized, allocation rule (5) is $h(\mathbf{e}_j^t\mathbb{H}^t\mathbb{W}\mathbb{H}\mathbf{D}_n)$, namely it depends on the jth component of $\mathbb{H}^t\mathbb{W}\mathbb{H}\mathbf{D}_n$.

Proposition 1 *Under the ECADE,* $\mathbf{D}_0 = \mathbf{0}_s$ *and* $\mathbf{D}_{n+1} = \mathbf{D}_n + (2\delta_{n+1} - 1)\mathbf{x}_{n+1}$ *a.s. for every* n*, so* $\{\mathbf{D}_n\}_{n\in\mathbb{N}}$ *is a Markov chain on* $\mathbb{Z}^s$ *with transition probability*

$$\Pr(\mathbf{D}_{n+1} = \mathbf{y}' \mid \mathbf{D}_n = \mathbf{y}) = \begin{cases} p_j h(\mathbf{e}_j^t \mathbb{H}^t \mathbb{W} \mathbb{H} \mathbf{y}), & \mathbf{y}' = \mathbf{y} + \mathbf{e}_j \\ p_j [1 - h(\mathbf{e}_j^t \mathbb{H}^t \mathbb{W} \mathbb{H} \mathbf{y})], & \mathbf{y}' = \mathbf{y} - \mathbf{e}_j, \end{cases} \tag{6}$$

for $j = 1, \ldots, s$*. Thus,* $\{\mathbf{D}_n\}_{n\in\mathbb{N}}$ *is irreducible and periodic, with period 2.*

Moreover, $\mathbf{b}_0 = \mathbf{0}_{q+1}$ *and* $\mathbf{b}_{n+1} = \mathbf{b}_n + (2\delta_{n+1} - 1)\mathbb{H}\mathbf{x}_{n+1}$ *a.s. for every* n*; thus,* $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ *is a Markov chain on* $\mathcal{B} = \{\mathbf{y} \in \mathbb{Z}^{q+1} : \mathbf{y} = \mathbb{H}\mathbf{x},$ *with* $\mathbf{x} \in \mathbb{Z}^s\}$ *with transition probability*

$$\Pr(\mathbf{b}_{n+1} = \mathbf{y}' \mid \mathbf{b}_n = \mathbf{y}) = \begin{cases} p_j h(\mathbf{e}_j^t \mathbb{H}^t \mathbb{W} \mathbf{y}), & \mathbf{y}' = \mathbf{y} + \mathbb{H}\mathbf{e}_j \\ p_j [1 - h(\mathbf{e}_j^t \mathbb{H}^t \mathbb{W} \mathbf{y})], & \mathbf{y}' = \mathbf{y} - \mathbb{H}\mathbf{e}_j, \end{cases}$$

for $j = 1, \ldots, s$*; thus,* $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ *is irreducible and periodic, with period 2.*

The main properties of the ECADE are given in the following Theorem.

Theorem 1 *For qualitative covariates, under the ECADE the irreducible Markov chain* $\{\mathbf{D}_n\}_{n\in\mathbb{N}}$ *is bounded in probability and thus* $\mathbf{D}_n = O_p(1)$*; therefore,* $\mathbf{b}_n = O_p(1)$ *and* $\|\mathbf{b}_n\|_{\mathbb{W}}^2 = O_p(1)$*, while the loss* ℓ_n *and the Mahalanobis distance* $\mathcal{M}_n$ *are asymptotically equivalent with an order of convergence* $o_p(1)$*.*

The proof can be found in Appendix A.2.

As shown in Example 1, the ECADE with $\mathbb{W} = \mathbb{P}_{n+1}^{-1}$ sequentially minimizes ℓ_n. In such a case $\mathbb{H}^t\mathbb{W}\mathbb{H}\mathbf{D}_n = [\mathbb{H}^t[\mathbb{H}\hat{\mathbf{P}}_{n+1}\mathbb{H}^t]^{-1}\mathbb{H}\hat{\mathbf{P}}_{n+1}]\hat{\mathbf{P}}_{n+1}^{-1}\mathbf{D}_n$, namely it is simply a projection of $\hat{\mathbf{P}}_{n+1}^{-1}\mathbf{D}_n = (D_{n1}/\hat{p}_{n+1,1}, \ldots, D_{ns}/\hat{p}_{n+1,s})^t$. Moreover, when the linear model is full $\mathbb{H}$ is non singular and $\mathbb{H}^t[\mathbb{H}\hat{\mathbf{P}}_{n+1}\mathbb{H}^t]^{-1}\mathbb{H}\hat{\mathbf{P}}_{n+1} = \mathbb{I}_s$ for every n; thus, adopting ECADE with the allocation function h_E, both choices $\mathbb{W} = \mathbb{P}_{n+1}^{-1}$ and $\mathbb{W} = \mathbb{P}_n^{-1}$ correspond to the same stratified randomization performed via Efron's BCD, since $h_E(\mathbf{e}_j^t\hat{\mathbf{P}}_{n+1}^{-1}\mathbf{D}_n) = h_E(\mathbf{e}_j^t\hat{\mathbf{P}}_n^{-1}\mathbf{D}_n) = h_E(\mathbf{e}_j^t\mathbf{D}_n)$. In this setting, at each stratum $\{D_{nj}\}_{n\in\mathbb{N}}$ is an irreducible and bounded in probability Markov chain on $\mathbb{Z}$ ($j = 1, \ldots, s$) and, from (2), $\ell_n = n^{-1}\mathbf{D}_n^t\mathbb{H}^t[\mathbb{H}\hat{\mathbf{P}}_n\mathbb{H}^t]^{-1}\mathbb{H}\mathbf{D}_n = n^{-1}\mathbf{D}_n^t\hat{\mathbf{P}}_n^{-1}\mathbf{D}_n = \sum_{j=1}^s D_{nj}^2/N_{nj} = o_p(1)$.

The next result provides the condition under which the imbalance vector and the stratum imbalance are still bounded in probability even if the weight matrix changes at each step of the sequential procedure.

Theorem 2 *Let* $\{\mathbb{W}_n\}_{n\in\mathbb{N}}$ *be a sequence of symmetric and positive-definite matrices such that* $\lim_{n\to\infty}\mathbb{W}_n = \mathbb{W}$ *a.s. (with* $\mathbb{W}$ *symmetric and positive-definite). Thus,* $\mathbf{D}_n = O_p(1)$ *and* $\mathbf{b}_n = O_p(1)$*, so that* $\|\mathbf{b}_n\|_{\mathbb{W}}^2 = O_p(1)$*, while* ℓ_n *and* $\mathcal{M}_n$ *are both* $o_p(1)$*.*

The proof is reported in Appendix A.3.

Remark 4 Hu and Hu (2012)'s procedure randomizes the assignments through $h_E(\cdot)$ by using as imbalance measure a weighted average of the global, marginal and stratum imbalances for

each profile. Thus, their procedure can be derived as a special case of the ECADE by setting $\mathbb{W}$ such that $\mathbb{H}^t\mathbb{W}\mathbb{H} = \mathbb{U}$, where $\mathbb{U}$ is defined in equation (6.2) of Hu and Hu (2012), provided that $\mathbb{U}$ is positive-definite. Clearly, this condition is not generally satisfied by any possible choice of their weights. For instance, taking into account the case of two binary covariates as in Section 3 in Hu and Hu (2012), $\mathbb{U}$ is a block centrosymmetric matrix of the form $\mathbb{U} = \mathbb{I}_2 \otimes \mathbb{U}_1 + \mathbb{J}_2 \otimes \mathbb{U}_2$, with

$$\mathbb{U}_1 = \begin{pmatrix} 1 & w_0 + w_{m1} \\ w_0 + w_{m1} & 1 \end{pmatrix}, \quad \text{and} \quad \mathbb{U}_2 = \begin{pmatrix} w_0 + w_{m2} & w_0 \\ w_0 & w_0 + w_{m2} \end{pmatrix},$$

where w_0 is the weight of the global balance, while w_{m1} and w_{m2} are the marginal weights of the two covariates. From the properties of centrosymmetric matrices, $\mathbb{U}$ is positive-definite if and only if the two matrices $\mathbb{U}_1 \pm \mathbb{J}_2\mathbb{U}_2$ are both positive-definite, namely if and only if $(1 - w_0)^2 > (w_{m1} - w_{m2})^2$ and $1 > w_0 + w_{m1} + w_{m2}$. Since the stratum weight $w_s = 1 - w_0 + w_{m1} + w_{m2}$, a necessary condition is that $w_s > 0$ and, assuming $w_{m1} = w_{m2}$ as in Corollary 3.1 in Hu and Hu (2012), $w_s > 0$ is also sufficient to guarantee the positive-definiteness of $\mathbb{U}$. Thus, our Theorem 1 relaxes the quite restrictive conditions (B) of Theorem 3.1 and (B') of Corollary 3.1 (and so (C) of Theorem 3.2 and (C') of Corollary 3.2) in Hu and Hu (2012). Essentially, with more than six covariates, condition (C) implies that $w_s \to 1$, namely Hu and Hu's rule corresponds to a stratified randomization performed via Efron's BCD.

3.2 Properties of the ECADE for quantitative and mixed covariates

Let now $\boldsymbol{Z}$ be a vector of quantitative covariate and assume that the (joint) density $\mathcal{L}(\boldsymbol{z})$ is defined on $\mathbb{R}^p$. The following Theorem shows that $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is bounded in probability.

Theorem 3 *For quantitative covariates, under the ECADE $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is a Markov chain on $\mathbb{Z} \times \mathbb{R}^q$ with*

$$\mathbf{b}_0 = \mathbf{0}_{q+1} \ \text{ and } \ \mathbf{b}_{n+1} = \mathbf{b}_n + (2\delta_{n+1} - 1)(1;\ \mathbf{f}(\boldsymbol{Z}_{n+1})^t)^t \ \text{ a.s. for every } n. \tag{7}$$

Moreover, if $\lim_{x\to\infty} h(x) = e \in (0, 1/2)$, then $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is bounded in probability, so that $\mathbf{b}_n = O_p(1)$ and therefore $\|\mathbf{b}_n\|^2_{\mathbb{W}} = O_p(1)$, while $\ell_n = o_p(1)$ and $\mathcal{M}_n = o_p(1)$. This result still holds even if $\mathbb{W}$ is replaced by a sequence $\{\mathbb{W}_n\}_{n\in\mathbb{N}}$ of symmetric and positive-definite matrices, provided that $\lim_{n\to\infty} \mathbb{W}_n = \mathbb{W}$ a.s.

For the proof see Appendix A.4.

From Theorem 3, we need to impose a mild restriction of the codomain of the allocation function to guarantee the boundedness in probability of the imbalance vector. While h_E directly guarantees boundedness, a general allocation function h can be suitably rescaled to satisfy this additional condition by setting, for instance, $h_e(x) = e + (1 - 2e)[1 - h(x)]$.

Moreover, Theorem 3 could be naturally extended to the case of mixed covariates, where p_1 qualitative factors are combined with p_2 quantitative ones (with $p_1 + p_2 = p$). In such a case, $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is a Markov chain on $\mathbb{Z}^{1+q_1} \times \mathbb{R}^{q_2}$, where q_1 and q_2 are the transformed dimensions via $\mathbf{f}(\cdot)$ of the two blocks of covariates (with $q_1 + q_2 = q$). Clearly, the transition in (7) still holds, so that $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is bounded in probability.

4 Finite sample properties and comparisons

In this section, an extensive simulation study is performed to illustrate the properties of the ECADE with $\mathbb{W} = \mathbb{P}_{n+1}^{-1}$ aimed at minimizing the loss (ECADE$_\ell$), also compared with other CA rules. Adopting model (1), we denote by M_E the model with only the main effects, namely with $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$, while M_I accounts for first order interactions. We stress that every categorical covariate with k levels is represented by a vector of $k-1$ dummies. The procedures are compared in terms of the simulated averages of ℓ_n and $\mathcal{M}_n$ in the case of qualitative covariates (Section 4.1) and quantitative/mixed factors (Section 4.2), while Section 4.3 points out the impact of the discretization of quantitative covariates. In all the scenarios, we perform 10000 Monte Carlo replications of the trial. For consistency in the comparisons, in this section the simulations of the ECADE have been run by setting $h = h_E$; since a value $\rho \in [0.8, 0.95]$ has been suggested for covariate adaptive randomization (Weir and Lees, 2003; Hu and Hu, 2012; Ma and Hu, 2013), in what follows we adopt $\rho = 0.85$.

4.1 Simulation results for qualitative covariates

The ECADE is compared to Atkinson's D_A-BCD (D_A), the minimization method of Pocock and Simon (1975) (PS), the Covariate Adaptive BCD (CABCD) suggested in Baldi Antognini and Zagoraiou (2011) with allocation function $F(x) = (x^5 + 1)^{-1}$ for $x \geq 1$, and the CA rule of Hu and Hu (2012) (HH). In the case of binary factors, we also consider the group sequential RR proposed by Zhou et al (2018): as suggested in their Supplementary Material, we set the total expected number of rerandomizations equal to 2000, while each group contains 50 experimental units. Concerning HH, as in Section 4.3 of Hu and Hu (2012) we fix $w_0 = w_s = 1/3$ and equal marginal weights. We assume $Z_1, \ldots, Z_p$ independent, each of them with equiprobable levels, inducing the uniform distribution $\mathbf{p} = s^{-1}\mathbf{1}_s$ over the strata.

The first study concerns $n = 400$ and $p = 3$ covariates $\boldsymbol{Z} = (Z_1, Z_2, Z_3)^t$, one binary and the others with three levels, generating $s = 18$ strata. The behavior of the simulated averages of ℓ_n and $\mathcal{M}_n$ is displayed in Figure 1 under models M_E and M_I.

For both ℓ_n and $\mathcal{M}_n$, the ECADE$_\ell$ represents the best choice, regardless of the chosen model. HH rule always guarantees good performances, although its loss is higher than the one of the ECADE$_\ell$. Given the small number of covariates, the CABCD performs quite good and under M_I it is the second best procedure for $n \geq 150$. Atkinson's rule presents a quite stable behavior: from around $n = 100$, its expected loss is very close to the corresponding theoretical value $(q+1)/5$. As expected, PS performs very well under M_E, whereas its performance deteriorates dramatically when the model complexity increases (the values of the expected loss are more than ten times higher than those of the ECADE$_\ell$ under M_I, with $n = 400$). As discussed in Theorem 1, the Mahalanobis distance is asymptotically equivalent to the loss, so $\mathcal{M}_n$ confirms the same conclusions of ℓ_n.

Let us now consider the case of $p = 10$ binary covariates, inducing 1024 strata. Figure 2 shows the simulated averages of ℓ_n and $\mathcal{M}_n$ under models M_E and M_I with $n = 1000$. Due to the large number of strata, the CABCD presents the worst performance with respect to the other CA rules: as expected, the stratified randomization

hardly evolves with a very small number of subjects in each stratum. Moreover, the inadequacy of PS in the presence of interactions among the covariates is confirmed and the expected loss of the D_A-BCD is quite large and does not vanish.

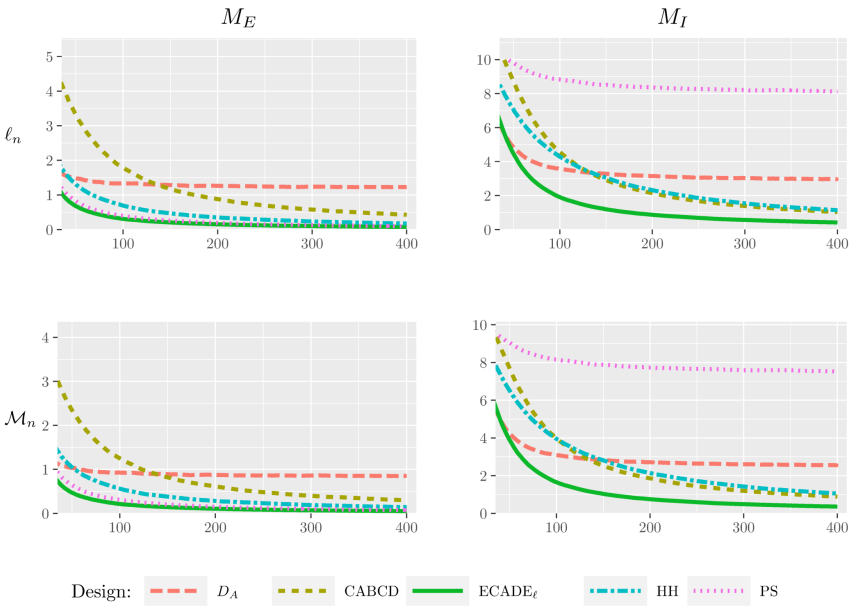

Figure 1: Simulated averages of ℓ_n and $\mathcal{M}_n$ for three categorical covariates under models M_E and M_I.

Under M_E, the expected loss of the ECADE$_\ell$, PS and HH present a quite similar behavior, but the ECADE$_\ell$ is still the best performing rule, showing values more than three times smaller than those of HH. Under M_I, the ECADE$_\ell$ results in a remarkable gain in terms of ℓ_n. For $n = 400$, the expected loss of HH is more than five times higher than that of the ECADE$_\ell$ and it goes up to ten times higher when $n = 1000$. As regards RR, the peculiarity of the plots stems from its group sequential nature. As discussed in the Supplementary Material of Zhou et al (2018), RR is known to perform well as long as some continuous covariates are present; indeed, our results show that RR may suffer from the presence of solely binary factors, especially either for small sample sizes and/or as the model complexity grows.

4.2 The case of quantitative or mixed covariates

For quantitative factors, we take into account a trial with $n = 600$, where the ECADE$_\ell$ is compared to the D_A-BCD and the sequential RR. In addition, we consider the kernel (KER) procedure introduced by Ma and Hu (2013), aimed at minimizing the difference between the covariate densities in the treatment arms. By using a kernel estimation method, the authors consider an imbalance measure which

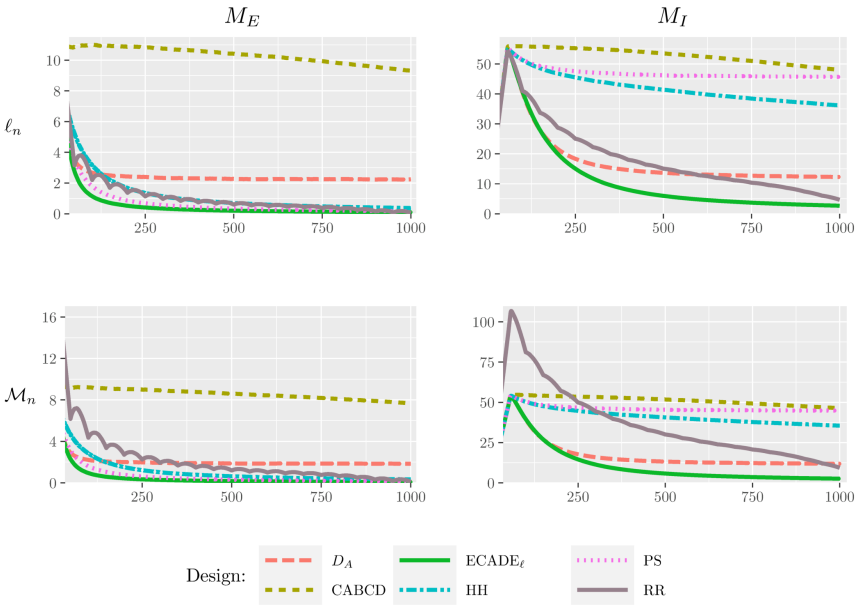

Figure 2: Simulated averages of ℓ_n and $\mathcal{M}_n$ for ten binary covariates under models M_E and M_I.

is a weighted average of the estimated distributional imbalances and the assignments are randomized via $h_E(\cdot)$. Following the authors' suggestion, we set equal weights for each factor and the covariates are re-scaled at each step on the basis of previously observed profiles, in order to have (asymptotically) the same unitary variance.

We first consider the case of $p = 3$ quantitative covariates $\boldsymbol{Z} = (Z_1, Z_2, Z_3)^t$, assumed to be independent and normally distributed with means $(3, 1, 2)^t$ and standard deviations $(2, 0.5, 1.5)^t$. The behavior of the competing CA rules is shown in Figure 3. Also for quantitative factors, the ECADE$_\ell$ tends to be the best procedure, regardless of the model and the chosen imbalance measure and RR performs similarly. KER has quite low values of the expected loss for M_E, while under M_I its performance dramatically worsen. As expected, under D_A-BCD the expected loss does not vanish. Again, $\mathcal{M}_n$ essentially presents the same behavior of ℓ_n.

By combining the previously discussed scenarios, we now take into account six covariates, three of them qualitative (one binary and the others having three levels) and three normally distributed factors. The simulated averages of ℓ_n and $\mathcal{M}_n$ are shown in Figure 4 with $n = 800$. The ECADE$_\ell$ is still the best procedure; the KER presents low values of both ℓ_n and $\mathcal{M}_n$ for the empty model, while its performance worsen in the presence of interactions among covariates. For the D_A-BCD, similar considerations of the case of three quantitative covariates still hold.

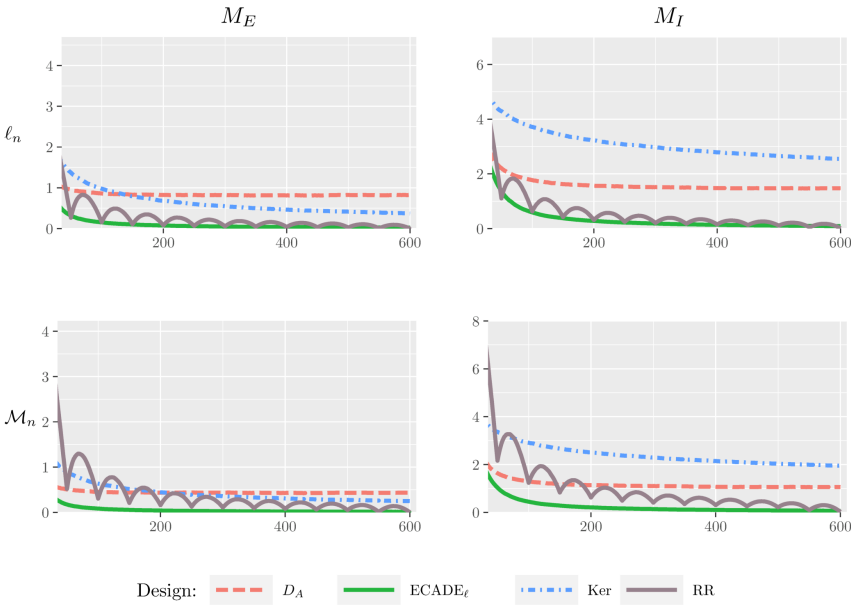

Figure 3: Simulated averages of ℓ_n and $\mathcal{M}_n$ for $p = 3$ quantitative covariates under models M_E and M_I.

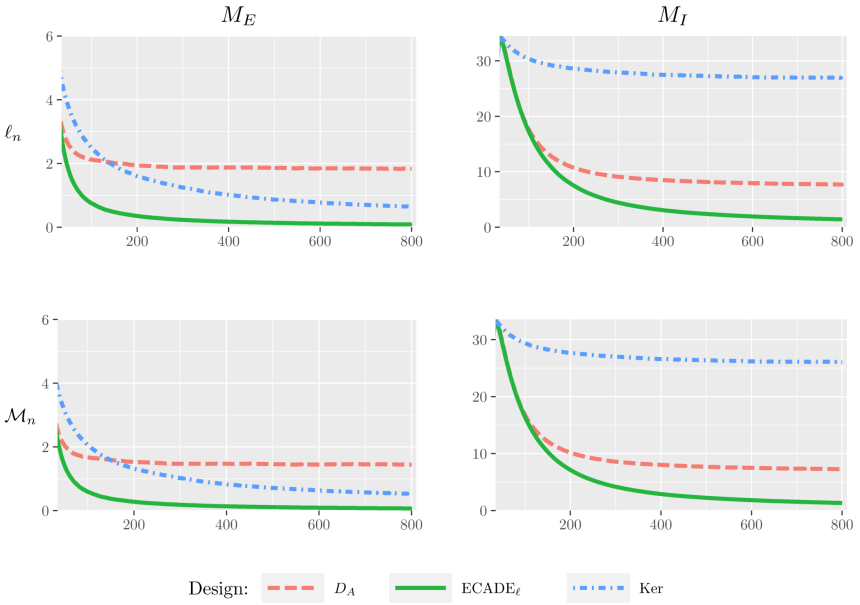

Figure 4: Simulated averages of ℓ_n and $\mathcal{M}_n$ for $p = 6$ mixed covariates under models M_E and M_I.

4.2.1 The effect of model misspecification

The relationship between covariates and the response variable is generally unknown in practice so it is important to assess the performance of our design in the case of model misspecification. The misspecification is accounted for by not including one or more covariates and/or not including one or more covariate effects. We run a simulation study in which we also considered comparison with other procedures. We first run the clinical trial adopting the true model and we compute ℓ_n and $\mathcal{M}_n$. Then, we simulate the same clinical trial adopting the misspecified model, which generates a different allocation vector, say $\boldsymbol{\delta}_n^M$ so a different imbalance vector, say $\mathbf{b}_n^M$, at the end of the trial. Then the loss under the misspecified model ℓ_n^M is computed adopting $\mathbf{b}_n^M$ as $n^{-1}(\mathbf{b}_n^M)^t\mathbb{P}_n^{-1}\mathbf{b}_n^M$ and the same principle is applied for computing $\mathcal{M}_n^M$. We consider a clinical experiment including six covariates: three binary and three continuous (with the same parameter specification of Section 4.2). In scenario I, the true model is such that $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$ (i.e. $p = 6$), while the misspecified one considers only four covariates: two binary and two continuous (so that $p = 4$). In scenario II the true model includes the main effects for all the six covariates, first order interactions among the three binary variables and quadratic effects for two of the three quantitative factors, while the adopted misspecified model is the one with $\mathbf{f}(\boldsymbol{Z}) = \boldsymbol{Z}$. The results are reported in Figure 5.

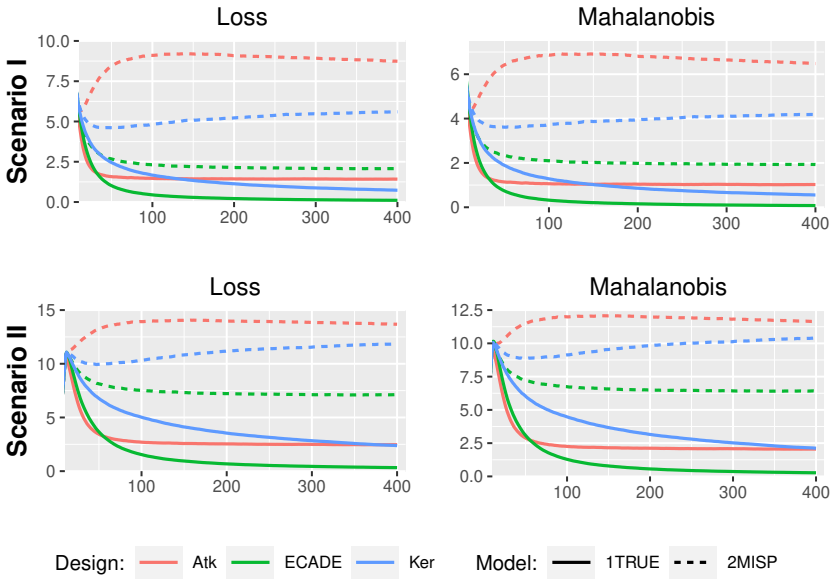

Figure 5: Simulated averages of ℓ_n, $\mathcal{M}_n$, ℓ_n^M and $\mathcal{M}_n^M$ for Scenario I and II. The behaviour of the considered procedures under the misspecified model is denoted by the superscript M.

As regards scenario I, the ECADE shows a milder inflation in the loss and Mahalanobis distance with respect to the other procedures, with values of ℓ_n and $\mathcal{M}_n$ that tend to be close to those of the D_A-BCD and Ker under a correct model specification. In Scenario II, the case of a more severe misspecification, the differences between ℓ_n and ℓ_n^M (as well as between $\mathcal{M}_n$ and $\mathcal{M}_n^M$) are more pronounced. However, the ECADE guarantees better performances also in this experimental setting, with values of ℓ_{400}^M equal to around half of those under the D_A design and to around 60% of those under the Ker procedure, respectively. Finally, it is worth noticing that under model misspecification, the loss and Mahalanobis distance of the Ker procedure are increasing with n and do not seem to reach a limiting value even for $n = 400$.

4.3 The effect of categorizing quantitative covariates

Due to the lack of CA rules aimed at achieving balance among continuous covariates, a common practice is to breakdown the factors into subcategories, which may strongly affect the loss of inferential precision (Atkinson, 2002; Ciolino et al, 2011; Ma and Hu, 2013). In particular, the impact of the discretization is related to the way in which the covariates are categorized (namely, the chosen cut-offs) and the knowledge of their generating distribution. In general, an inflated loss can be avoided only when the covariate distribution is a-priori known, so it is customary to choose the median for dichotomizing the factors.

To show the impact of the discretization of covariates in terms of loss, also compared with CA rules that account naturally on mixed factors, we perform a simulation study under the previously discussed scenario with $p = 3$ normal covariates (having means and standard deviations given by $(3, 1, 2)^t$ and $(2, 0.5, 1.5)^t$, respectively) and, to account for non-symmetrical distributions, we also consider the left-truncated normal with truncation in 1. These variables have been dichotomized by selecting the correct median or an incorrect value (that is, $1.5\cdot$median) as the cut-off. In summary, we consider the following experimental settings: normal distribution and correct median (N-c), normal distribution and incorrect median (N-inc), truncated normal and correct median (TN-c) and truncated normal with incorrect median (TN-inc). We wish to stress that, in the real practice, the covariate distribution is unknown (in terms of both, symmetry and median), especially in the sequential recruitment where the initial information is poor.

Table 1 reports the simulated averages of ℓ_n for $n = 200$ and 400 under M_E and model M_F with $\mathbf{f}(Z_1, Z_2, Z_3) = (Z_1, Z_2, Z_3, Z_1 \cdot Z_2, Z_1 \cdot Z_3, Z_2 \cdot Z_3, Z_1 \cdot Z_2 \cdot Z_3)^t$, where ECADE$_\ell(d)$ and $D_A(d)$ are implemented with the categorized factors, while ECADE$_\ell$ and D_A keep the continuity of the covariates. First of all, the effect of non-symmetric distributions is extremely moderate, provided that the cut-offs exactly coincide with the median. Under M_E, the effect of an incorrect choice of the threshold is notable and becomes even more remarkable when the distribution is not symmetric (only Atkinson's procedure is quite invariant). In general, adopting the ECADE$_\ell$ and D_A-BCD with the original quantitative factors induces a smaller loss than those of the corresponding dichotomized versions and this effect is stronger as the model complexity grows, where the impact of discretization becomes critical for some CA rules. Under M_F, excluding PS which is strongly inadequate in this case

regardless of the nature of the covariates, the loss for the listed procedures is always higher than the one achieved by the ECADE$_\ell$ even if the cut-offs are correctly chosen. When the median is not correctly specified, the loss inflation highly grows. For example, under the normal distribution with $n = 400$, the CABCD's loss is 4.5 times higher, while HH's loss is four times bigger with respect to its value in N-c. If compared to the ECADE$_\ell$, for $n = 400$ and N-inc the loss of the CABCD and that of HH are around 8 times bigger. In the case of TN-inc these values reach up to nearly ten. Finally, note that the ECADE$_\ell(d)$ is still the best choice for discretized factors. These results are even more pronounced when p increases, as well as when the discretization involves more than two categories, leading to multiple improper choices of the cut-offs, as further simulations have shown (see also Lauzon et al (2020)).

Table 1: Simulated averages of ℓ_n for $p = 3$ factors under M_E (no interactions) and M_F. For PS, CABCD, HH, $D_A(d)$ and ECADE$_\ell(d)$, the covariates have been dichotomized.

	M_E				M_F			
	N-c		N-inc		N-c		N-inc	
ℓ_n	200	400	200	400	200	400	200	400
PS	0.08	0.04	0.15	0.08	4.14	4.01	4.10	4.11
CABCD	0.26	0.13	0.46	0.25	0.52	0.26	1.87	1.17
HH	0.14	0.07	0.22	0.11	0.58	0.28	1.74	1.13
$D_A(d)$	0.83	0.82	0.83	0.82	1.71	1.66	2.09	1.90
D_A	0.83	0.82	0.83	0.82	1.57	1.49	1.57	1.49
ECADE$_\ell(d)$	0.07	0.04	0.10	0.05	0.34	0.17	1.23	0.75
ECADE$_\ell$	0.07	0.04	0.07	0.04	0.28	0.14	0.28	0.14
	TN-c		TN-inc		TN-c		TN-inc	
	200	400	200	400	200	400	200	400
PS	0.09	0.04	0.28	0.14	4.12	3.99	3.36	3.61
CABCD	0.27	0.13	0.64	0.37	0.60	0.29	2.42	1.93
HH	0.14	0.07	0.34	0.18	0.65	0.32	2.37	2
$D_A(d)$	0.83	0.81	0.86	0.82	1.70	1.64	2.26	2.12
D_A	0.83	0.81	0.83	0.82	1.57	1.50	1.57	1.50
ECADE$_\ell(d)$	0.08	0.04	0.17	0.08	0.38	0.19	1.90	1.53
ECADE$_\ell$	0.08	0.04	0.08	0.04	0.28	0.14	0.39	0.19

Distribution of covariates and chosen cut-off: N-c (normal distribution and correct median); N-inc (normal distribution and incorrect median); TN-c (truncated normal and correct median); TN-inc (truncated normal with incorrect median).

5 Redesign of NIDA-CENIC-P1S1 clinical trial

This section is dedicated to the application of our proposed methodology to redesign a real trial: we use the clinical data of National Institute on Drug Abuse (NIDA-CENIC-P1S1) project (Donny et al, 2015), freely available at National Institute of Health - NIDA Data Share Website - https://datashare.nida.nih.gov/. The aim of the

study was to investigate the impact of nicotine using spectrum cigarettes (P1S1) and to assess the relationship between very low nicotine cigarettes and cigarette/tobacco use, nicotine/tobacco exposure and dependence along with other health-related behaviors. The original database consisted of baseline and demographic information of $n = 839$ patients: we choose eleven covariates, five quantitative (three discrete and two continuous) and six qualitative, that have few missing values and are regarded as most likely to be associated with the research object. The qualitative covariates are subdivided into dummy variables according to their different levels: this led to a total of 34 covariates considered. The missing values are imputed by sampling from the observed values. In what follows, in order to highlight the flexibility of the new proposal, two different allocation functions $h(\cdot)$ for the ECADE minimizing the loss of information are considered:

- Efron's h_E with $\rho = 0.85$, simply indicated with ECADE$_\ell$
- $h_e(x) = e + (1 - 2e)[1 - \Phi(x)]$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution, with $e = 0.1$ (ECADE$_{\ell_1}$) and $e = 0.2$ (ECADE$_{\ell_2}$).

Adopting M_E, the performance of the ECADE are compared with those of the D_A-BCD and the Kernel (Ker) procedure with $\rho = 0.85$ by simulating 10000 replications.

The behavior of ℓ_n and $\mathcal{M}_n$ are shown in Figure 6, while Table 2 summarizes the values of the imbalance measures along with their variability.

Table 2: Simulated averages of ℓ_n and $\mathcal{M}_n$ with their standard error (in brackets) for NIDA-CENIC-P1S1 data.

	ℓ_n		$\mathcal{M}_n$	
ECADE$_\ell$	2.42	(1.00)	2.31	(0.96)
ECADE$_{\ell_1}$	2.01	(0.87)	1.94	(0.84)
ECADE$_{\ell_2}$	3.06	(1.21)	2.97	(1.19)
D_A	7.44	(1.79)	6.85	(1.69)
Ker	15.89	(5.88)	15.83	(5.87)

The results confirm the excellent ability of ECADE$_\ell$ to balance the experimental groups: values of both ℓ_n and $\mathcal{M}_n$ are about one third of those of the D_A-BCD (e.g. 2.42 vs 7.34) and about one seventh of the Kernel ones (e.g. 2.42 vs 15.89), which seems to suffer the presence of a high number of covariates. It should be noted that the improvement with respect to the other procedures is also evident for small/moderate sample sizes and increases as the numbers of patients grows. For what concerns our proposal, ECADE$_{\ell_1}$ enforces a higher balance also for smaller sample sizes, while Efron's allocation function shows an intermediate performance. Moreover, regardless of the allocation function adopted, the new proposal shows the lowest variability in the estimates.

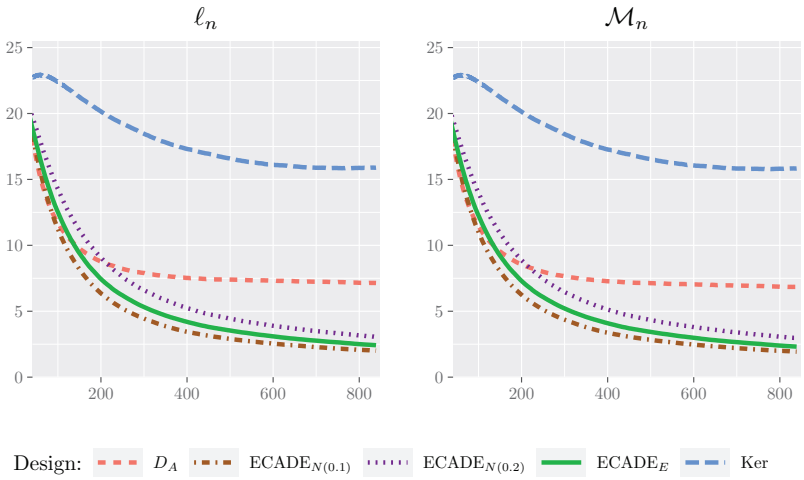

Figure 6: Simulated averages of ℓ_n and $\mathcal{M}_n$ for NIDA-CENIC-P1S1 data.

6 Conclusions

For valid comparisons, the sequential allocation of treatments should be randomized with the aim of obtaining a trial as balanced as possible. Due to the recent applications of big data, clinical trials with biomarkers and medical genomic, the need of procedures that are high-order balanced among a set of pre-specified covariates, often very numerous and of mixed nature, is growing. Nowadays, the most popular and commonly implemented methods are stratification and marginal procedures. However, both of them (as well as Hu and Hu's rule) can be applied solely to categorical covariates, that implies an arbitrary discretization of quantitative factors. As pointed out in Section 4.3, this discretization process could make balance unreliable, also strongly damaging the inferential precision, especially when the covariate distribution is a-priori unknown. Therefore, HH rule, stratified and marginal procedures are recommended only for truly qualitative factors. Alternative rules, such as Atkinson's D_A-BCD and the design of Ma and Hu (2013), can deal with continuous covariates, but do not guarantee a suitable degree of balance. Whereas, sequential RR performs very good in the case of continuous factors, but it may suffer when all the covariates are binary or when some of them are polythomic, since they must be dichotomized (see Section 4.1 and the Supplementary Material of Zhou et al (2018)).

In this paper we propose a new family of covariate-adaptive randomization rules minimizing a suitable weighted norm of the imbalance vector. We prove that the ECADE is high-order balanced, regardless of both the number of chosen factors as well as their nature, allowing to avoid discretization and the above-mentioned implications. In particular, the ECADE$_\ell$ (minimizing the loss of inferential precision) guarantees excellent performances with respect to all the other considered rules. Although its strength lies in balancing quantitative or mixed covariates more efficiently, the ECADE performs very well even in the presence of many categorical factors (or

levels) as well as for interactions/higher order effects, in contrast to the behavior of stratified randomization which cannot evolve properly in such situations. Moreover, the ECADE$_\ell$ guarantees a substantial gain of estimation precision also with respect to the D_A-BCD and Pocock and Simon's minimization method. Furthermore, even if ECADE$_\ell$ has the same order of convergence to balance of Hu and Hu's rule, the former guarantees higher levels of balance regardless of the model and the chosen imbalance measure for any sample size.

Despite the wide use of covariate adaptive designs and their implications in terms of estimation precision (Baldi Antognini and Zagoraiou, 2017), concerns on the validity of conventional inferential procedures have been raised by many authors. Starting from Shao et al (2010), several papers discussed the theoretical properties of hypothesis testing for CA randomization procedures for discrete/discretized covariates. In particular, for CA designs with overall and marginal imbalances bounded in probability, Ma et al (2015) showed that the usual tests for the linear model are always correct in terms of type I error provided that the inferential analysis is performed by including all the covariates used in the randomization process; otherwise these tests become conservative when some factors are omitted in the analysis. It would be desirable to extend such results in the presence of continuous and mixed covariates and we hope that this paper will contribute to provide a basis for future developments in this direction.

An interesting possible extension that we will consider as future research is the case of several treatments. Obviously, the chosen optimality criteria should be based on the contrasts instead of the main effects, leading to possible reformulation of the imbalance vector. Thus, the ECADE could be generalized to the multi-arm case by assigning each treatment proportionally to the corresponding gain in terms of the predicted imbalance.

Appendix A Proofs

A.1 Optimality of balance for estimating/testing the treatment difference $\mu_A - \mu_B$

For simplicity of notation in this Appendix we omit the subscript n. After n allocations, let $\hat{\boldsymbol{\gamma}}$ be the LSE of $\boldsymbol{\gamma} = (\mu_A, \mu_B, \boldsymbol{\beta})^t$ and $a^t = (1;\ -1;\ \mathbf{0}_q^t)$, then $\text{var}(a^t\hat{\boldsymbol{\gamma}}) = \sigma^2 a^t (n\mathbb{M})^{-1} a$, where

$$\mathbb{M} = \frac{1}{n}\begin{pmatrix} n_A & 0 & \boldsymbol{\delta}^t\mathbb{F} \\ 0 & n_B & (\mathbf{1}-\boldsymbol{\delta})^t\mathbb{F} \\ \mathbb{F}^t\boldsymbol{\delta} & \mathbb{F}^t(\mathbf{1}-\boldsymbol{\delta}) & \mathbb{F}^t\mathbb{F} \end{pmatrix}.$$

Let $\mathbf{x}^t = \mathbf{1}^t\mathbb{F}$ and $\mathbf{y}^t = (2\boldsymbol{\delta} - \mathbf{1})^t\mathbb{F}$, then $n\mathbb{M} = \left(\begin{array}{c|c} \text{diag}\,(n_A, n_B) & \mathbb{D}^t \\ \hline \mathbb{D} & n\mathbb{A} \end{array}\right)$, with $\mathbb{D} = 2^{-1}\,(\mathbf{x}+\mathbf{y} \mid \mathbf{x}-\mathbf{y})$. By letting $\mathbb{T} = \text{diag}\,(n_A, n_B) - \mathbb{D}^t(n\mathbb{A})^{-1}\mathbb{D}$, then

$a^t(n\mathbb{M})^{-1}a = (1,-1)\mathbb{T}^{-1}\begin{pmatrix}1\\-1\end{pmatrix}$, where

$$\mathbb{T} = \left(\begin{array}{c|c} n_A - (4n)^{-1}(\mathbf{x}+\mathbf{y})^t\mathbb{A}^{-1}(\mathbf{x}+\mathbf{y}) & -(4n)^{-1}(\mathbf{x}+\mathbf{y})^t\mathbb{A}^{-1}(\mathbf{x}-\mathbf{y}) \\ \hline -(4n)^{-1}(\mathbf{x}-\mathbf{y})^t\mathbb{A}^{-1}(\mathbf{x}+\mathbf{y}) & n_B - (4n)^{-1}(\mathbf{x}-\mathbf{y})^t\mathbb{A}^{-1}(\mathbf{x}-\mathbf{y}) \end{array}\right).$$

After some algebra, $\det\mathbb{T} = (n - \mathbf{x}^t(n\mathbb{A})^{-1}\mathbf{x})(n-\ell)/4$, so that

$$a^t(n\mathbb{M})^{-1}a = \frac{n - \mathbf{x}^t(n\mathbb{A})^{-1}\mathbf{x}}{(n - \mathbf{x}^t(n\mathbb{A})^{-1}\mathbf{x})(n-\ell)/4} = \frac{4}{n-\ell} = \frac{4}{n}\left(1 - \frac{\ell}{n}\right)^{-1}. \quad \text{(A1)}$$

Taking into account hypothesis testing, under well-known regularity conditions $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{q+2}, \sigma^2\mathbb{M}^{-1})$, so that $\sqrt{n}a^t(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} \mathrm{N}(0, \sigma^2\|a\|^2_{\mathbb{M}^{-1}})$. Assuming σ^2 known, the classical Wald statistic is $W = n\hat{\boldsymbol{\gamma}}^t a[\sigma^2\|a\|^2_{\mathbb{M}^{-1}}]^{-1}a^t\hat{\boldsymbol{\gamma}} = n(\hat{\mu}_A - \hat{\mu}_B)^2[\sigma\|a\|_{\mathbb{M}^{-1}}]^{-2}$. Under H_0, $W \xrightarrow{d} \chi^2_1$, namely it converges to a (central) χ^2 with 1 degree of freedom (dof); whereas, under the alternative W converges to a non-central χ^2_1 with non-centrality parameter $n(\mu_A - \mu_B)^2[\sigma\|a\|_{\mathbb{M}^{-1}}]^{-2}$. From (A1), $a^t\mathbb{M}^{-1}a = 4\,(1-\ell/n)^{-1}$, so the non-centrality parameter is equal to $(2\sigma)^{-2}n(\mu_A - \mu_B)^2(1-\ell/n)$. For fixed dof the non-central χ^2 is stochastically increasing in the non-centrality parameter. Thus, for every sample size the power is an increasing function of it and is maximized when $\ell = 0$.

A.2 Proof of Theorem 1

Following Theorem 4.5 of Meyn and Tweedie (1992), a Markov chain $\{\mathbf{X}_n\}_{n\in\mathbb{N}}$ on a general state-space $\mathbb{X}$ is bounded in probability if i) $\mathbf{X}_n$ is a T-chain and ii) $\mathbf{X}_n$ satisfies a positive drift condition, namely there exists a norm-like function $V : \mathbb{X} \longrightarrow \mathbb{R}^+$ such that, for some $\varepsilon > 0$ and a compact set $\mathcal{C} \in \mathcal{B}(\mathbb{X})$ (where $\mathcal{B}(\mathbb{X})$ is the Borel sigma algebra), we have

$$(D) \quad \Delta V(\mathbf{X}_n) := \mathbb{E}[V(\mathbf{X}_{n+1}) \mid \mathbf{X}_n] - V(\mathbf{X}_n) \leq -\varepsilon, \qquad \mathbf{X}_n \in \mathcal{C}^c.$$

Since a Markov chain on a countable state space is always a T-chain (Meyn and Tweedie (1992), p. 548) in what follows we just need to show that condition (D) is satisfied by $\mathbf{D}_n$ and $\mathbf{b}_n$. Let $\mathbb{C} = \mathbb{H}^t\mathbb{W}\mathbb{H}$, then $\mathbb{C} = (c_{ij})_{i,j=1,\ldots,s}$ is symmetric and positive-definite, since $\mathbb{H}$ is a full (row) rank matrix. Moreover, by letting $\tilde{\mathbf{D}}_n = \mathbb{H}^t\mathbb{W}\mathbb{H}\mathbf{D}_n$, the linear transformation $\tilde{\mathbf{D}}_n = \mathbb{C}\mathbf{D}_n$ is an isomorphism and therefore the behavior of the Markov chain $\{\mathbf{D}_n\}_{n\in\mathbb{N}}$ is equivalent to the one of $\{\tilde{\mathbf{D}}_n\}_{n\in\mathbb{N}}$ (although defined in a proper transformed space) and their roles in the proof could be naturally exchanged. We show that condition (D) holds by setting $V(\mathbf{D}_n) = \mathbf{D}_n^t\mathbb{C}\mathbf{D}_n$ and $\mathcal{C} = \{\mathbf{D}_n : \|\mathbb{C}\mathbf{D}_n\|_\infty \leq \kappa\}$, where $\kappa > 0$. Note that $\|\mathbb{C}\mathbf{D}_n\|_\infty$ is still a norm of the vector $\mathbf{D}_n$, since $\mathbb{C}$ is invertible (namely the corresponding linear transformation is injective). Due to the isomorphism, the compact set could be analogously expressed by $\mathcal{C} = \{\tilde{\mathbf{D}}_n : \|\tilde{\mathbf{D}}_n\|_\infty \leq \kappa\}$, so

$\mathcal{C}^c = \{\tilde{\mathbf{D}}_n : \max_{j=1,\dots,s} |\tilde{D}_{nj}| > \kappa\}$. From (6),

$$\Delta V(\mathbf{D}_n) = \sum_{j=1}^{s} \mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j] \Pr(\mathbf{x}_{n+1} = \mathbf{e}_j), \quad \text{(A2)}$$

where

$$\begin{aligned}&\mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j] = \\ &\Pr(\delta_{n+1} = 1 \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j)\, \mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j, \delta_{n+1} = 1] + \\ &\Pr(\delta_{n+1} = 0 \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j)\, \mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j, \delta_{n+1} = 0].\end{aligned} \quad \text{(A3)}$$

Moreover,

$$\begin{aligned}&\mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \,|\mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j, \delta_{n+1} = 1] = \\ &\qquad = (\mathbf{D}_n + \mathbf{e}_j)^t \mathbb{C}(\mathbf{D}_n + \mathbf{e}_j) - \mathbf{D}_n^t \mathbb{C}\mathbf{D}_n = c_{jj} + 2\mathbf{e}_j^t \mathbb{C}\mathbf{D}_n\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \,|\mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j, \delta_{n+1} = 0] = \\ &\qquad = (\mathbf{D}_n - \mathbf{e}_j)^t \mathbb{C}(\mathbf{D}_n - \mathbf{e}_j) - \mathbf{D}_n^t \mathbb{C}\mathbf{D}_n = c_{jj} - 2\mathbf{e}_j^t \mathbb{C}\mathbf{D}_n,\end{aligned}$$

where $c_{jj} = \mathbf{e}_j^t \mathbb{C} \mathbf{e}_j > 0$. Thus, recalling that $\tilde{\mathbf{D}}_n = \mathbb{C}\mathbf{D}_n$, equation (A3) becomes

$$\begin{aligned}\mathbb{E}[V(\mathbf{D}_{n+1}) - V(\mathbf{D}_n) \mid \mathbf{D}_n, \mathbf{x}_{n+1} = \mathbf{e}_j] &= h(\mathbf{e}_j^t \tilde{\mathbf{D}}_n)[c_{jj} + 2\mathbf{e}_j^t \tilde{\mathbf{D}}_n] + \\ + [1 - h(\mathbf{e}_j^t \tilde{\mathbf{D}}_n)][c_{jj} - 2\mathbf{e}_j^t \tilde{\mathbf{D}}_n] &= 4\tilde{D}_{nj} \left[h(\tilde{D}_{nj}) - \frac{1}{2}\right] + c_{jj}.\end{aligned} \quad \text{(A4)}$$

Therefore, from (A2) and (A4), $\Delta V(\mathbf{D}_n) = 4\sum_{j=1}^{s} \tilde{D}_{nj}[h(\tilde{D}_{nj}) - 1/2]p_j + \sum_{j=1}^{s} c_{jj} p_j$, where the first term is always non-positive since, if $\tilde{D}_{nj} \geq 0$ then $h(\tilde{D}_{nj}) \leq 1/2$ and when $\tilde{D}_{nj} < 0$ then $h(\tilde{D}_{nj}) > 1/2$. Condition (D) is equivalent to

$$\sum_{j=1}^{s} |\tilde{D}_{nj}|[h(-|\tilde{D}_{nj}|) - 1/2]p_j \geq (\varepsilon + \sum_{j=1}^{s} c_{jj} p_j)/4. \quad \text{(A5)}$$

By the definition of $\mathcal{C}^c$, there exists at least one stratum $\tilde{j}$ such that $|\tilde{D}_{n\tilde{j}}| > \kappa$ and therefore $\sum_{j=1}^{s} |\tilde{D}_{nj}|[h(-|\tilde{D}_{nj}|) - 1/2]p_j \geq \kappa[h(-|\kappa|) - 1/2]p_{\tilde{j}}$, which is an increasing function of κ since $h(-|\kappa|) > 1/2$ and $p_{\tilde{j}} > 0$. Thus, for every $\mathbf{D}_n \in \mathcal{C}^c$ condition (D) is verified, since the RHS of (A5) is bounded.

In addition, since $\mathbf{b}_n = \mathbb{H}\mathbf{D}_n$, then $\mathbf{b}_n = O_p(1)$ as a bounded linear combination of $\mathbf{D}_n$; thus, from (2) and (4), $\ell_n = o_p(1)$, since $\|\mathbf{b}_n\|^2_{\mathbb{P}_n^{-1}}$ is asymptotic equivalent to $\|\mathbf{b}_n\|^2_{\mathbb{P}^{-1}} = O_p(1)$ (recalling that $\mathbb{P}_n - \mathbb{P} = o_{a.s.}(1)$). At the same time $\mathbf{b}_n = O_p(1)$, then $n_A \bar{\mathbf{f}}_n^A - n_B \bar{\mathbf{f}}_n^B = O_p(1)$ and $n^{-1} D_n = o_p(1)$; by letting $\pi_n = n_A/n$,

then $\pi_n - 1/2 = o_p(1)$ and $\|n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}} = O_p(1)$, so that $n^{-1}\|n_A\bar{\mathbf{f}}_n^A - n_B\bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}} = n\|\pi_n\bar{\mathbf{f}}_n^A - (1-\pi_n)\bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}} = o_p(1)$. Since $\pi_n - 1/2 = o_p(1)$ and $\mathbb{A}_n - \mathbb{A} = o_{a.s.}(1)$, then $n\|\pi_n\bar{\mathbf{f}}_n^A - (1-\pi_n)\bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}_n^{-1}}$ is asymptotic equivalent to $4^{-1}n\|\bar{\mathbf{f}}_n^A - \bar{\mathbf{f}}_n^B\|^2_{\mathbb{A}^{-1}} = o_p(1)$, namely $\mathcal{M}_n = o_p(1)$. Thus, under the ECADE, ℓ_n is asymptotic equivalent to $\mathcal{M}_n$. By the same arguments ℓ_n and $\mathcal{M}_n$ are asymptotic equivalent for every CA rule under which $\pi_n - 1/2 = o_p(1)$.

A.3 Proof of Theorem 2

Let $\mathbb{C}_n = \mathbb{H}^t\mathbb{W}_n\mathbb{H}$, since $\mathbb{W}_n \longrightarrow \mathbb{W}$ a.s., then $\mathbb{C}_n \longrightarrow \mathbb{C} = \mathbb{H}^t\mathbb{W}\mathbb{H}$ a.s. from Slutshy's theorem. Consider the function $W(\mathbf{D}_n) = \mathbf{D}_n^t\mathbb{C}_n\mathbf{D}_n = U(\mathbf{D}_n) + V(\mathbf{D}_n)$, where $U(\mathbf{D}_n) = \mathbf{D}_n^t(\mathbb{C}_n - \mathbb{C})\mathbf{D}_n$, while $V(\cdot)$ and the compact set $\mathcal{C}$ are the same as in the proof of Theorem 1 in Section A.2. Thus we show that condition (D) is satisfied by $\Delta W(\mathbf{D}_n) = \mathbb{E}[W(\mathbf{D}_{n+1}) - W(\mathbf{D}_n) \mid \mathbf{D}_n] = \Delta U(\mathbf{D}_n) + \Delta V(\mathbf{D}_n)$. Since $\mathbb{C}_n - \mathbb{C} = o_{a.s.}(1)$, then $\Delta U(\mathbf{D}_n) = \mathbb{E}[\mathbf{D}_{n+1}^t(\mathbb{C}_{n+1} - \mathbb{C})\mathbf{D}_{n+1} - \mathbf{D}_n^t(\mathbb{C}_n - \mathbb{C})\mathbf{D}_n \mid \mathbf{D}_n]$ tends to be negligible for a sufficiently large n. Moreover, from Section A.2, $\Delta V(\mathbf{D}_n)$ satisfies condition (D); thus the Markov chain induced by $\mathbb{W}_n$ is asymptotically equivalent to the one corresponding to $\mathbb{W}$, and therefore $\mathbf{D}_n = O_p(1)$.

A.4 Proof of Theorem 3

Under the ECADE with quantitative factors $\mathbf{b}_{n+1} = \mathbf{b}_n + (2\delta_{n+1} - 1)(1;\ \mathbf{f}(\boldsymbol{Z}_{n+1})^t)^t$ a.s. for every n (with $\mathbf{b}_0 = \mathbf{0}_{q+1}$) and therefore $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is Markov chain on $\mathbb{X} = \mathbb{Z} \times \mathbb{R}^q$ with one step transition kernel

$$
\begin{aligned}
P(\mathbf{x}, A) &= \Pr(\mathbf{b}_{n+1} \in A \mid \mathbf{b}_n = \mathbf{x}) \\
&= \int \Pr(\mathbf{b}_{n+1} \in A \mid \mathbf{b}_n = \mathbf{x}, \mathbf{Z}_{n+1} = \boldsymbol{z})\mathcal{L}(\boldsymbol{z})d\boldsymbol{z} \\
&= \int \Big\{ h\left((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{x}\right) I\{\mathbf{x} + (1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in A\} + \\
&\qquad + \left[1 - h\left((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{x}\right)\right] I\{\mathbf{x} - (1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in A\} \Big\} \mathcal{L}(\boldsymbol{z})d\boldsymbol{z}.
\end{aligned}
\tag{A6}
$$

Following Theorem 4.5 of Meyn and Tweedie (1992), we firstly show that $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is a T-chain and then we prove that condition (D) is satisfied. As regards the T-chain property, we need to show that there exists a sampling distribution a and a substochastic transition kernel $T(\mathbf{x}, \cdot)$ such that for any $A \in \mathcal{B}(\mathbb{X})$, $K_\alpha(\mathbf{x}, A) = \sum_{i=1}^{\infty} P^i(\mathbf{x}, A)\alpha(i) \geq T(\mathbf{x}, A)$, where $T(\cdot, A)$ is a lower semicontinuous (LSC) function with $T(\mathbf{x}, \mathbb{X}) > 0$ for all $\mathbf{x} \in \mathbb{X}$. By taking $\alpha(1) = 1$ and 0 otherwise, then $K_\alpha(\mathbf{x}, A) = P(\mathbf{x}, A)$; if we set $T(\mathbf{x}, A) = e \int I\{\mathbf{x} + (1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in A\}\mathcal{L}(\boldsymbol{z})d\boldsymbol{z}$, then $P(\mathbf{x}, A) \geq T(\mathbf{x}, A)$, recalling that $h(x) \geq e$ for any $x \in \mathbb{R}$. Since the indicator function of any open set is LSC, $T(\mathbf{x}, A)$ is always LSC. Indeed, if A is an open subset then

$$\begin{aligned}\liminf_{\mathbf{y}\to\mathbf{x}} T(\mathbf{y},A) &= e \liminf_{\mathbf{y}\to\mathbf{x}} \int I\{\mathbf{y}+(1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in A\}\mathcal{L}(\boldsymbol{z})d\boldsymbol{z} \\ &\geq e \int \liminf_{\mathbf{y}\to\mathbf{x}} I\{\mathbf{y}+(1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in A\}\mathcal{L}(\boldsymbol{z})d\boldsymbol{z} \qquad \text{(A7)} \\ &\geq e \int I\{\mathbf{x}+(1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in A\}\mathcal{L}(\boldsymbol{z})d\boldsymbol{z} = T(\mathbf{x},A).\end{aligned}$$

Moreover, (A7) holds even if A is not open; indeed, $T(\mathbf{x},A)$ does not change since the closure of A has zero Lebesgue measure. Finally notice that $T(\mathbf{x},\mathbb{X})>0$ since $\mathbf{x}+(1;\ \mathbf{f}(\boldsymbol{z})^t)^t \in \mathbb{Z}\times\mathbb{R}^q$ a.s. Thus, $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is a T-chain.

We now show that condition (D) is satisfied by choosing $V(\mathbf{b}_n)=\mathbf{b}_n^t\mathbb{W}\mathbf{b}_n$. The one-step drift is

$$\begin{aligned}\Delta V(\mathbf{b}_n) &= \mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n] = \mathbb{E}\left[\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b})\mid \mathbf{b}_n=\mathbf{b},\mathbf{Z}_{n+1}=\boldsymbol{z}]\right] \\ &= \int \mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z}]\mathcal{L}(\boldsymbol{z})d\boldsymbol{z},\end{aligned}$$

where the inner expectation is

$$\begin{aligned}&\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z}] = \\ &\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z},\delta_{n+1}=1]h((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n)+ \\ &\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z},\delta_{n+1}=0][1-h((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n)].\end{aligned}$$

Since $\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z},\delta_{n+1}=1] = 2(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n + (1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}(1;\ \mathbf{f}(\boldsymbol{z})^t)^t$ and $\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z},\delta_{n+1}=0] = -2(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n + (1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}(1;\ \mathbf{f}(\boldsymbol{z})^t)^t$, then $\mathbb{E}[V(\mathbf{b}_{n+1})-V(\mathbf{b}_n)\mid \mathbf{b}_n,\mathbf{Z}_{n+1}=\boldsymbol{z}] = 4(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\left[h\left((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\right)-\frac{1}{2}\right] + (1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}(1;\ \mathbf{f}(\boldsymbol{z})^t)^t$, so that

$$\begin{aligned}\Delta V(\mathbf{b}_n) = 4\int (1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\left[h\left((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\right)-\frac{1}{2}\right]\mathcal{L}(\boldsymbol{z})d\boldsymbol{z}+ \\ + \int (1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}(1;\ \mathbf{f}(\boldsymbol{z})^t)^t\mathcal{L}(\boldsymbol{z})d\boldsymbol{z}.\end{aligned}$$

Notice that $(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\left[h\left((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\right)-1/2\right] \leq 0$, since $(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\left[h\left((1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n\right)-1/2\right]=0$ if and only if $(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n=0$ and it is negative otherwise. To verify condition (D), we need to show that, for a compact set $\mathcal{C}$, $\Delta V(\mathbf{b}_n)\leq -\varepsilon$, namely

$$\begin{aligned}\int |(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n|\left[h\left(-|(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n|\right)-\frac{1}{2}\right]\mathcal{L}(\boldsymbol{z})d\boldsymbol{z} \geq \\ \frac{\varepsilon+\int (1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}(1;\ \mathbf{f}(\boldsymbol{z})^t)^t\mathcal{L}(\boldsymbol{z})d\boldsymbol{z}}{4}, \quad \text{on } \mathcal{C}^c.\end{aligned} \qquad \text{(A8)}$$

Let $\mathcal{Z}^* = \{\boldsymbol{z} : |(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n| > 0\} \subset \mathbb{R}^p$, then $\Pr(\boldsymbol{z} \in \mathcal{Z}^*) > 0$ and the LHS of (A8) becomes

$$\int_{\mathcal{Z}^*} |(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n| \left[h\left(-|(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n|\right) - \frac{1}{2}\right] \mathcal{L}(\boldsymbol{z})d\boldsymbol{z}.$$

Let $\mathcal{C} = \{\mathbf{b}_n : \max|(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n| \leq \kappa, \boldsymbol{z} \in \mathcal{Z}^*\}$ be the compact set (in $\mathcal{Z}^*$ the linear transformation $(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n$ is injective and corresponds to an induced norm of $\mathbf{b}_n$), for every $\mathbf{b}_n \in \mathcal{C}^c$,

$$\int_{\mathcal{Z}^*} |(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n| \left[h\left(-|(1;\ \mathbf{f}(\boldsymbol{z})^t)\mathbb{W}\mathbf{b}_n|\right) - \frac{1}{2}\right] \mathcal{L}(\boldsymbol{z})d\boldsymbol{z} > \kappa \left[h(-|\kappa|) - \frac{1}{2}\right] \Pr(\boldsymbol{z} \in \mathcal{Z}^*);$$

so condition (D) is verified since $\kappa\left[h(-|\kappa|) - \frac{1}{2}\right]\Pr(\boldsymbol{z} \in \mathcal{Z}^*)$ increases in κ, while the RHS of (A8) is bounded. Finally, the last statement follows from Theorem 2.

Under the mixed scenario, through the usual factorization of the joint distribution of mixed random variables, $\mathcal{L}(\boldsymbol{z})$ in (A6) will be substituted by the product of the joint probability distribution of p_1 qualitative covariates and the conditional density function of the remaining p_2 factors. Thus, the T-chain property is preserved and, under the same choice of the function V and the compact set $\mathcal{C}$, the positive drift condition is also satisfied, so $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ is bounded in probability.

References

Atkinson AC (1982) Optimum biased coin designs for sequential clinical trials with prognostic factors. Biometrika 69:61–67

Atkinson AC (2002) The comparison of designs for sequential clinical trials with covariate information. Journal of the Royal Statistical Society Series A 165:349–373

Baldi Antognini A, Giovagnoli A (2015) Adaptive Designs for Sequential Treatment Allocation. Chapman & Hall/CRC Biostatistics

Baldi Antognini A, Zagoraiou M (2011) The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. Biometrika 98:519–535

Baldi Antognini A, Zagoraiou M (2012) Multi-objective optimal designs in comparative clinical trials with covariates: the reinforced doubly-adaptive biased coin design. The Annals of Statistics 40:1315–1345

Baldi Antognini A, Zagoraiou M (2017) Estimation accuracy under covariate-adaptive randomization procedures. Electronic Journal of Statistics 11:1180–1206

Begg CB, Iglewicz B (1980) A treatment allocation procedure for sequential clinical trials. Biometrics 36:81–90

Ciolino J, Zhao W, Martin R, et al (2011) Quantifying the cost in power of ignoring continuous covariate imbalances in clinical trial randomization. Contemporary Clinical Trials 32(2):250 – 259

Donny EC, Denlinger RL, Tidey JW, et al (2015) Randomized trial of reduced-nicotine standards for cigarettes. New England Journal of Medicine 373(14):1340–1349

Efron B (1971) Forcing sequential experiments to be balanced. Biometrika 58:403–417

Heritier S, Gebski V, Pillai A (2005) Dynamic balancing randomization in controlled clinical trials. Statistics in Medicine 24:3729–3741

Hu Y, Hu F (2012) Asymptotic properties of covariate-adaptive randomization. The Annals of Statistics 40:1794–1815

Lauzon SD, Ramakrishnan V, Nietert PJ, et al (2020) Statistical properties of minimal sufficient balance and minimization as methods for controlling baseline covariate imbalance at the design stage of sequential clinical trials. Statistics in Medicine 39:2506–2517

Ma W, Hu F, Zhang L (2015) Testing hypotheses of covariate-adaptive randomized clinical trials. Journal of the American Statistical Association 110(510):669–680

Ma Z, Hu F (2013) Balancing continuous covariates based on kernel densities. Contemporary Clinical Trilas 34(2):262–269

Meyn SP, Tweedie RL (1992) Stability of markovian processes i: criteria for discrete time markov chains. Advances in Applied Probability 24(3):542–574

Morgan KL, Rubin D (2012) Rerandomization to improve covariate balance in experiments. The Annals of Statistics 40(2):1263 – 1282

Pocock SJ, Simon R (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31:103–115

Rosenberger WF, Lachin JL (2002) Randomization in Clinical Trials: Theory and Practice. John Wiley & Sons, New York

Shao J, Yu X, Zhong B (2010) A theory for testing hypotheses under covariate-adaptive randomization. Biometrika 97:347–360

Smith RL (1984a) Properties of biased coin designs in sequential clinical trials. The Annals of Statistics 12:1018–1034

Smith RL (1984b) Sequential treatment allocation using biased coin designs. Journal of the Royal Statistical Society Series B 46:519–543

Taves DR (1974) Minimization: a new method of assigning patients to treatment and control groups. Journal of Clinical Pharmacy and Therapeutics 15:443–453

Wei LJ (1978) The adaptive biased coin design for sequential experiments. The Annals of Statistics 6:92–100

Weir CJ, Lees KR (2003) Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. Statistics in medicine 22(5):705–726

Zhou Q, Ernst PA, Morgan KL, et al (2018) Sequential Rerandomization. Biometrika 105(3):745 – 752