Universidad
Pública de Navarra
Nafarroako
Unibertsitate Publikoa

Probabilistic analysis of adaptive experimental designs

Arkaitz Galbete

Supervisors:

José Antonio Moler

Fernando Plo

Introducción

En este trabajo se persiguen dos objetivos generales:

a) El estudio de diseños experimentales adaptados a la respuesta y la propuesta de un nuevo diseño competitivo en propiedades tanto éticas como inferenciales respecto a los ya existentes.

b) El análisis de la inferencia basada en la aleatorización en el marco de diseños adaptados a la respuesta.

Las herramientas utilizadas en este trabajo son fundamentalmente probabilísticas, cadenas de Markov en tiempo discreto tanto homogéneas como no homogéneas, teoría de martingalas y leyes límite de procesos. También se han utilizado técnicas de inferencia paramétrica y no paramétrica.

Los diseños estadísticos en los que se centra esta tesis son los ensayos clínicos. Un ensayo clínico es, en su versión más habitual, un diseño que consiste en comparar el comportamiento de dos tratamientos (dos fármacos o dos protocolos de actuación ante una enfermedad) con objeto de establecer la superioridad de uno de ellos respecto al otro. Los ensayos clínicos se diferencian de la mayoría de diseños de experimentos por dos razones principales. Por un lado, las unidades experimentales son los pacientes, con lo que el experimento adquiere un componente ético que no es propio de otros experimentos estadísticos. Este hecho implica una fuerte regulación y control del ensayo por parte de las autoridades competentes. Por otro lado, la llegada de los pacientes puede ser secuencial, pudiendo pasar periodos importantes de tiempo entre las diferentes llegadas. Debido a esta llegada secuencial, la información se va acumulando y está disponible mientras se lleva a cabo el experimento. Parece entonces natural el uso de esta información para introducir alguna mejora en el experimento. Ante la llegada de un paciente, este es asignado a uno de los dos tratamientos mediante una regla de asignación aleatoria que utiliza la información pasada del ensayo. Estos ensayos se denominan ensayos clínicos adaptativos. Si la asignación se hace utilizando las asignaciones pasadas junto con las

respuestas pasadas, el diseño se denomina adaptado a la respuesta. El trabajo se centra en este tipo de diseños, siendo el núcleo de esta tesis la presentación y estudio de diversos aspectos de un diseño de ensayo clínico adaptado a la respuesta de fase III, llamado diseño de urna de Klein.

En el capítulo 1 se hace una presentación del tema de la tesis así como una revisión bibliográfica de los tres temas principales que se trabajan en ella, los diseños adaptativos, con especial énfasis en los diseños adaptados a la respuesta, la inferencia basada en la aleatorización y el uso de covariables en diseños adaptativos. En el capítulo 2 se presentan tres diseños adaptados a la respuesta diferentes, basados en el diseño de urna de Ehrenfest, para después efectuar un análisis de sus propiedades y elegir uno para un estudio más profundo, debido a sus prometedoras propiedades. Este diseño se llamará diseño de urna de Klein, debido a que se basa en un proceso de urna ya estudiado por Martin J. Klein dentro del campo de la física. En este mismo capítulo se hace un estudio exacto y asintótico del los procesos estocásticos asociados al diseño, así como un análisis de propiedades más concretas de los ensayos clínicos, como sesgo de selección, sesgo accidental o un estudio de potencia y de las propiedades éticas del diseño.

Los capítulos 3 y 4 se dedican a la inferencia basada en la aleatorización. Debido a la particularidad del proceso de selección de pacientes de un ensayo clínico, en el que la obtención de la muestra de pacientes depende de diversos factores difícilmente controlables, la asunción de un modelo poblacional puede ser puesta en duda, con lo que se opta por asumir un modelo de aleatorización. En este marco de trabajo inferencial, las respuestas de los pacientes se consideran fijas y no se asume ningún modelo probabilístico para ellas. Por tanto, la aleatoriedad solo descansa en el proceso de asignación de pacientes, llamado aleatorización, y que da base a la inferencia. Bajo este modelo, se presenta un estadístico de test lineal de rangos para respuestas dicotómicas. En el capítulo 3 se obtiene un algoritmo recursivo que permite el cálculo de la distribución exacta del estadístico y que hace factible el cálculo de los p-valores en un tiempo razonable. Este algoritmo es válido

para una amplia gama de diseños adaptados a la respuesta en la que está incluido el diseño de urna de Klein. Asimismo, el algoritmo es adaptado para un conocido diseño que no pertenece a esta familia. Utilizando el algoritmo, se completa un estudio de simulación para la potencia del test con diferentes diseños y estadísticos. El capítulo 4 se centra en el estudio asintótico del estadístico para la inferencia basada en la aleatorización con el diseño de urna de Klein. Las propiedades probabilísticas de la urna de Klein bajo este marco de trabajo cambian sustancialmente, con lo que muchas de las propiedades analizadas en el capítulo 2 necesitan ser adaptadas. Utilizando estas nuevas propiedades se obtiene la distribución asintótica del estadístico del test.

Por último, en el capítulo 5 se introducen las covariables en los diseños adaptados a la respuesta, particularmente con el diseño de urna de Klein. Se presenta una versión estratificada de dicho diseño, analizando las convergencias de los procesos y estimadores básicos. Por otro lado, se propone la utilización de un modelo lineal generalizado para determinar la relación de la respuesta con los tratamientos y covariables. Por la dependencia de los errores de los diseños adaptados, la teoría clásica, basada en la incorrelación de los errores, no es aplicable, con lo que es necesario demostrar una ley fuerte y un teorema central para los estimadores de máxima cuasi-verosimilitud. Finalmente, se realiza un estudio de simulación comparando la inferencia basada en aleatorización con la inferencia paramétrica utilizando un modelo lineal generalizado.

Introduction

This thesis has two general goals:

a) The study of experimental response-adaptive designs and the proposal of a new design, with good ethical and inferential properties and that it is competitive with respect to the existing ones.

b) The analysis of randomization based inference for response-adaptive designs.

In this thesis, the main tools are probabilistic, as time-homogeneous and non-homogeneous Markov chains, martingale theory and limit laws. Some concepts about nonparametric and parametric inference have also been used.

This dissertation is focused on clinical trials. Commonly, a clinical trial is a statistical design to compare the behavior of two treatments (two drugs or two intervention protocols) in order to establish the superiority of one with respect to the other. Two main aspects distinguish clinical trials from the majority of experimental designs. On the one hand, the experimental units are patients, so the experiment is concerned about ethics, and this does not happen in other statistical experiments. This fact implies a strong regulation and control of the trial by the law and the relevant authorities. On the other hand, the arrival of the patients can be sequential with long enough interarrival times. Therefore, the information is accrued and it is available during the experiment. Then, it seems natural to use this information in order to improve the properties of the design. When a patient arrives, he or she is assigned to a treatment via a random allocation rule that takes into account the past information. These designs are called adaptive designs. When the allocation is made using the past allocations and past responses of the patients, it is called response-adaptive. This work is focused on these designs and the core of the thesis is the presentation and study of different features of a response-adaptive phase III clinical trial design, which is called Klein urn design.

In Chapter 1 the state of the art is reviewed, focusing on the three main addressed topics, which are: adaptive designs, particularly response-adaptive designs, randomization

based inference and the use of covariates in response-adaptive designs. In Chapter 2 three response-adaptive designs are presented, based on the Ehrenfest urn design. After doing an analysis of their properties, the one with the best properties is chosen for a thorough study. This design is named Klein urn design, since it was studied as an urn process by Martin J. Klein in the field of physics. An exact and asymptotic study of some stochastic processes related to the design are also completed in this chapter. Finally, some characteristic properties of clinical trials are analyzed when the Klein urn design is used to allocate patients, as selection bias, accidental bias, power study with parametric inference and ethical properties of the design.

Chapter 3 and Chapter 4 are focused on randomization based inference. In a clinical trial, the sampling process differs from the random sampling and the assumption of a population model is in doubt, so a randomization model is assumed. In this inferential framework, the patients' responses are treated as deterministic and no probabilistic model is assumed for them. Then, the only randomness is due to the random allocation of patients, called randomization. Under this model, a linear rank statistic for dichotomous responses is presented. In Chapter 3, a recursive algorithm is obtained to compute the exact distribution of the test statistic. This algorithm makes feasible the computation of the p-values and is valid for a wide family of response-adaptive designs, where the Klein urn design is included. Moreover, the algorithm is adapted also for a well-known design out of this family. With this algorithm, a simulation study is completed for the power of the test using different test statistics. In Chapter 4, an asymptotic study of a test statistic for randomization based inference is made when randomization is made with the Klein urn. The stochastic properties of the Klein urn change substantially so the properties analyzed in Chapter 2 must be revisited. Using these new properties, the asymptotic distribution of the test statistic is obtained.

Finally, in Chapter 5, the introduction of covariates in response-adaptive designs, specially in the Klein urn design, is studied. A stratified version of the Klein urn design is

presented and the convergence of the basic processes and estimators is proven. Besides this, a generalized linear model is used to determine the relation of the responses with the allocations and covariates. Due to the dependence relations generated when adaptive designs are used, the theory based on incorrelation of the error terms cannot be applied. Then, the consistency and asymptotic normality of the quasi maximum likelihood estimators must be proven. To conclude, a simulation study is carried out comparing randomization based inference and parametric inference using a generalized linear model.

Contents

Chapter 1

Background and literature review

A clinical trial is an experiment performed on human beings in order to evaluate the superiority of the treatment under study against another treatment or a placebo. A standard trial has different phases with different goals. In phase I, the potential toxicity of the new treatment is measured. Phase II studies the effectiveness of different doses. In phase III a comparison study is done to determine the superiority of the new treatment and finally, in phase IV a follow-up of the patients is done, see [58] for a detailed explanation. The particular structure of the experiment makes clinical trials quite different from the usual statistical experiments, leading to a complicate but challenging statistical design and analysis. We are going to focus on phase III.

The general framework in this thesis is a design with two treatments, say 1 and 2, with immediate and binary responses. We assume that patients arrive sequentially to the experiment. When a patient arrives, he or she is assigned with a randomized allocation rule to a treatment. The corresponding treatment is applied and the response of the patient, success or failure, is observed. Let δ_n be the indicator random variable of treatment 1 assignment; $\delta_n = 1$ if the treatment 1 is applied to the n-th patient and $\delta_n = 0$ otherwise. The indicator of treatment 2 assignment is $1-\delta_n$. $N_{n,1} = \sum_{i=1}^{n} \delta_i$ is the number of patients assigned to treatment 1 up to the n-th patient. Let $\mathbf{Z}_n = (Z_{n,1}, Z_{n,2})$ be the vector of potential responses of the n-th patient, considering that, for each n, only one is observed.

Unless otherwise stated, we assume the following parametric model. $Z_{n,1}$ and $Z_{n,2}$ are binary variables, taking value 1 with probability p_1 and p_2, respectively, if the response of the patient is a success, and being 0 with probability $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$, respectively, if the response of the patient is a failure.

1.1 Adaptive Randomization

One of the main advances in the design of clinical trials has been the use of the information that the experiment is giving in order to achieve several specific or global goals. These designs are called adaptive designs. When the accrued information is used to perform treatment allocations, we have an adaptive allocation rule. Depending on the previous information that we use to allocate the next patient, several kinds of adaptive allocation rules can be distinguished. Following the approach in [13], when the previous patient allocations are used to make the next allocation, we are in a treatment-adaptive design. If the responses of the patients are also included in the past information, the rule is called response-adaptive. Some covariates, that carry additional information about the patient, as gender or age, can be taken into account in the allocation rule for each patient n. This information will be denoted $\boldsymbol{H}_n$. If we use the covariate information of the past patients, the design is called covariate-adaptive. Finally, if we use all the past information, i.e. allocations, covariates and responses of the patients and, also, the covariate of the present patient before his allocation, we are in a covariate-adjusted response-adaptive (CARA) design, see [78]. The σ-algebra containing the information used to assign the next patient is denoted $\mathcal{F}_n$. This information is different depending on the kind of design. Let

$$\begin{aligned} \mathcal{G}_n &= \sigma(\delta_1, \ldots, \delta_n) && \text{previous allocations,} \\ \mathcal{Z}_n &= \sigma(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) && \text{previous responses,} \\ \mathcal{H}_n &= \sigma(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_n) && \text{previous covariates.} \end{aligned}$$

Then, if we are in a treatment-adaptive design, $\mathcal{F}_n = \mathcal{G}_n$, if the design is response-

adaptive, $\mathcal{F}_n = \sigma(\mathcal{G}_n, \mathcal{Z}_n)$. In case of covariate-adaptive designs, $\mathcal{F}_n = \sigma(\mathcal{G}_n, \mathcal{H}_n)$ and finally, in CARA designs, $\mathcal{F}_n = \sigma(\mathcal{G}_n, \mathcal{Z}_n, \mathcal{H}_{n+1})$.

The use of adaptive designs has an effect in the properties of the usual estimators, due to the inclusion of dependence relations in the processes. Some partial results are known. For instance, in [61] the consistency and asymptotic normality of the maximum likelihood estimators for adaptive designs are ensured, under some conditions that emulate those of the independence setting. A similar result is obtained in [43] for the least square estimators. A more general result was developed in [51] and in [11]. They proved that in the case of independent responses of the patients, if these responses are obtained in an adaptive allocation process, they remain independent. Then, if the number of allocations to each treatment tends to infinity, the estimators maintain the consistency and asymptotic normality of the estimators in the independence setting.

Now, we make a presentation of several adaptive designs that will be used in the contents of the thesis.

1.1.1 Complete Randomization

Complete randomization, (CR), is a non-adaptive randomization rule, where each treatment has a probability of 1/2 to be assigned to a patient. When the patient arrives, we throw a fair coin and the result of the coin determines the allocation. As we have independent throws of a balanced coin, it is well known that

$$\frac{N_{n,1}}{n} \to \frac{1}{2}\,, \quad a.s.$$

and

$$2\sqrt{n}\left(\frac{N_{n,1}}{n} - \frac{1}{2}\right) \to N(0,1)\,, \quad [D],$$

where $a.s.$ and D denote, respectively, almost sure convergence and convergence in distribution.

One of the main drawbacks of complete randomization is a non-negligible probability

of having an imbalance between treatment allocations, as stated in [63], Chapter 3. Those imbalances affect the statistical precision of the analysis and decrease the power of the statistical test. Due to this problem, some improvements were introduced in the allocation rules, in order to mitigate the probability of large imbalance.

1.1.2 Treatment-adaptive randomization

Treatment-adaptive designs, sometimes also called in the literature adaptive designs, use the previous treatment allocations to assign the next patient with the goal of increasing the probability of balance. For instance, they give more probability to the treatment that has been assigned fewer times. As a consequence, for all treatment-adaptive designs, we have that

$$\frac{N_{n,1}}{n} \to \frac{1}{2}\,, \quad a.s.$$

Introduced in [19], Efron's design was one of the first adaptive designs. When a patient arrives, a biased coin is thrown. The probability of treatment 1 allocation changes depending on the imbalance between the number of treatment assignments. It is $p > 1/2$ if this treatment has been assigned fewer times in the past and $1-p$ if the treatment has been assigned more times. When both treatments have been equally assigned, a fair coin is thrown. Denoting the imbalance of assignments as $D_n = N_{n,1} - N_{n,2}$, then,

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = \begin{cases} p & \text{if } D_n < 0\,, \\ 1-p & \text{if } D_n > 0\,, \\ \frac{1}{2} & \text{if } D_n = 0\,. \end{cases} \tag{1.1}$$

The main goal of Efron's design is to reduce the probability of imbalance between treatment allocations. Efron suggested $p = 2/3$ as an optimal choice for the value of p. However, the bias towards the under-represented treatment is the same regardless of the size of the imbalance. In [70], Smith presented a design where the probability of assigning treatment 1 to the n-th patient is given by the next family of functions:

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = \frac{(N_{n,2})^\rho}{(N_{n,1})^\rho + (N_{n,2})^\rho}\,, \qquad \rho \geq 0\,. \tag{1.2}$$

As in Efron's design, the probabilities are skewed to the under-represented treatment, but in Smith's design, the greater is the imbalance in the allocations, the bigger is the probability of assigning the under-represented treatment. Thus, imbalance is corrected faster. The value of the parameter ρ determines the degree of skewness of the allocation function. Very large values could produce almost deterministic designs, Smith suggested the value $\rho = 5$, which combines a fast convergence to balance and an acceptable level of randomness.

A generalization of Smith's designs, called ABCD designs, was proposed in [7]. They use a family of functions $F(D_n) : \mathbb{Z} \to [0, 1]$, decreasing and symmetric in the sense that $F(x) = 1 - F(1 - x)$, as an allocation rule.

Wei's urn design was presented in [73] with the idea of obtaining a faster balance between treatment allocations. The probabilistic tool used to randomize the allocation is an urn. Initially, w balls of type 1 and w balls of type 2 are in the urn. Type 1 balls are associated to treatment 1 and type 2 balls to treatment 2. When a ball is drawn, the corresponding treatment is applied and the ball is returned to the urn along with $\alpha > 0$ balls of the same type and $\beta > 0$ balls of the other type. Taking $\beta > \alpha$ the number of balls of both types tends to be balanced. The replacement rule is summarized in a matrix, called replacement matrix. The elements of the matrix are the number of added balls, where rows indicate the extraction of the different type of balls and columns indicate the effect in the different type of balls in the urn. The replacement matrix of Wei's urn design is

$$\begin{array}{lc} & \begin{array}{cc} add & add \\ type\ 1 & type\ 2 \\ balls & balls \\ \downarrow & \downarrow \end{array} \\ \begin{array}{l} \text{Get type 1 ball} \ \to \\ \text{Get type 2 ball} \ \to \end{array} & \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \end{array}.$$

Then, the probability of assigning treatment 1 is

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = \frac{w + \alpha N_{n,1} + \beta N_{n,2}}{2w + n\,(\alpha + \beta)}\,. \tag{1.3}$$

The Ehrenfest urn design was proposed for the first time in the context of clinical trials in [12]. As in Wei's urn design, initially the urn has w balls of each type, that is, $(W_{0,1}, W_{0,2}) = (w, w)$, where $W_{0,i}$ represents the initial number of type i balls in the urn. When a new patient arrives, a ball is drawn from the urn, the patient is allocated to the treatment associated with the type of ball and a ball of the other type is added to the urn. Let $W_{n,1}$ and $W_{n,2}$ be the number of balls of type 1 and type 2, respectively, after the n-th replacement. Observe that the total number of balls in the urn remains constant along the process, $W_{n,1} + W_{n,2} = 2w$. The replacement matrix for the Ehrenfest urn design has the following form,

$$\begin{array}{lc} & \begin{array}{cc} \textit{type } 1 & \textit{type } 2 \\ \downarrow & \downarrow \end{array} \\ \begin{array}{l} \text{Get type 1 ball} \ \rightarrow \\ \text{Get type 2 ball} \ \rightarrow \end{array} & \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}.$$

The probability of assigning treatment 1 is,

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = \frac{w - N_{n,1} + N_{n,2}}{2w}\,. \tag{1.4}$$

1.1.3 Response-adaptive randomization

As stated before, in response-adaptive randomization, each assignment is made using the previous treatment allocations and the previous responses of the patients. The main purpose of introducing the response information in the allocation rule is to improve the ethical aspect of the experiment. As we are collecting the responses obtained from each treatment application, we can promote the allocation to the treatment which is performing better and, so, skew the assignments towards this treatment by increasing its allocation probability. This improvement in the ethics of the design makes more difficult the study and

analysis of the design, because the dependence relationships among the different processes of the design become more complex. There is a wide collection of response-adaptive designs presented in the literature, good reviews can be found in [33] and [3]. Based on the procedure to assign patients, two types of designs can be differentiated. On the one hand, there are designs based on urn models. As some of the previous designs, such as Wei's urn design (1.3) and the Ehrenfest urn design, (1.4), the patient's allocation is made by drawing a ball from the urn, and the urn is updated taking into account the response of the patient, see, for instance [45]. On the other hand, there are biased coin designs, in which the probabilities of allocation of the different treatments are given by a function. A target quantity for the proportion of treatment 1 allocations, ρ, is fixed beforehand, and the probability of allocation varies in order to ensure the convergence to this target. Usually, the target depends on some parameters related to the responses of the patients. These quantities are commonly unknown, so they have to be estimated sequentially, and due to this fact these designs are also called *sequential estimation procedures.*

The play-the-winner rule (PTW) was one of the first response-adaptive designs presented in the literature and also one of the most studied, see [75]. In this model, initially we have an urn with ω balls of each treatment. When a patient arrives, a ball is extracted, the corresponding treatment is applied and the response is observed. If the response is a success, we return the ball along with β balls of the same treatment and α balls of the other treatment, being $\beta > \alpha$. In case of failure, we return the ball along with α balls of the same treatment and β of the other one. With this replacement policy a success is rewarded adding more balls of the applied treatment, and the failure is punished by adding more balls of the other treatment. The replacement matrix is

$$\begin{pmatrix} \beta Z_{n,1} + \alpha(1 - Z_{n,1}) & \alpha Z_{n,1} + \beta(1 - Z_{n,1}) \\ \alpha Z_{n,2} + \beta(1 - Z_{n,2}) & \beta Z_{n,2} + \alpha(1 - Z_{n,2}) \end{pmatrix}. \tag{1.5}$$

The probability of allocating treatment 1 can be expressed as

$$P(\delta_{n+1}=1|\mathcal{F}_n)=\frac{w+\alpha(F_{n,1}+S_{n,2})+\beta(S_{n,1}+F_{n,2})}{2w+n\,(\alpha+\beta)}\,,$$

where $S_{n,i}$ and $F_{n,i}$ are, respectively, the number of successes and failures in treatment i up to and including the n-th patient. The limit allocation of treatment 1 is:

$$\frac{N_{n,1}}{n}\to\frac{\alpha p_2+\beta q_2}{\alpha(p_1+p_2)+\beta(q_1+q_2)}\,,\quad a.s.$$

Usually, α is taken as 0 and β is taken as 1, and then

$$\frac{N_{n,1}}{n}\to\frac{q_2}{q_1+q_2}=\frac{\frac{1}{q_1}}{\frac{1}{q_1}+\frac{1}{q_2}}\,,\quad a.s.$$

If $p_1+p_2<3/2$ a central limit theorem is also obtained in [67],

$$\sqrt{n}\left(\frac{N_{n,1}}{n}-\frac{q_2}{q_1+q_2}\right)\to N(0,\sigma^2),\quad [D]$$

with

$$\sigma^2=\frac{q_1q_2(5-2(q_1+q_2))}{(2(q_1+q_2)-1)(q_1+q_2)^2}\,.$$

In [39] a modification of the play-the-winner rule is presented. It is called drop-the-loser (DTL) rule. The allocation of patients is done with an urn. This urn contains three types of balls, type 1 and type 2 corresponding to each treatment, and the so called immigration balls, which are denoted as type 0 balls. When a ball is extracted, if it is an immigration ball, we do not apply any treatment and the ball is returned along with one ball of each of the other types. When a ball of type 1 or 2 is extracted, the corresponding treatment is applied and the response observed. If it is a success, the ball is returned, with no change in the urn composition. Otherwise, if it is a failure, the ball is not replaced. This model punishes the failure by dropping balls, so the immigration balls are necessary to ensure that the urn does not get empty. The replacement matrix is

$$\begin{array}{l}\text{Get type 0 ball}\ \to\\ \text{Get type 1 ball}\ \to\\ \text{Get type 2 ball}\ \to\end{array}\begin{pmatrix}0 & 1 & 1\\ 0 & Z_{n,1}-1 & 0\\ 0 & 0 & Z_{n,2}-1\end{pmatrix}. \tag{1.6}$$

The limit of the proportion of treatment 1 allocations is the same as in the PTW rule, with $\alpha = 0$,

$$\frac{N_{n,1}}{n} \to \frac{q_2}{q_1 + q_2} \quad a.s.$$

A central limit theorem also holds for any p_1 and p_2,

$$\sqrt{n}\left(\frac{N_{n,1}}{n} - \frac{q_2}{q_1 + q_2}\right) \to N(0, \sigma^2), \quad [D],$$

with

$$\sigma^2 = \frac{q_1 q_2 (p_1 + p_2)}{(q_1 + q_2)^3}.$$

This model was generalized in [77]. The main advantage of this generalization is that after an immigration ball extraction, the number of type 1 and 2 balls added could be different, allowing different asymptotic allocation ratios.

The randomly reinforced urn design (RRU), presented in [47] in a more general setting, is based on an urn model which rewards a treatment when it is a success, tending to assign the best treatment with probability 1. Initially, we have an urn with $W_{0,1}$ balls of type 1 and $W_{0,2}$ balls of type 2. When a patient arrives, a ball is extracted, the corresponding treatment is applied and the response, $Z_{n,i}$, is observed. Then we put back the ball along with $Z_{n,i}$ balls of the same type. The replacement matrix stands as

$$\begin{pmatrix} Z_{n,1} & 0 \\ 0 & Z_{n,2} \end{pmatrix}. \tag{1.7}$$

The proportion of treatment 1 assignments satisfies,

$$\frac{N_{n,1}}{n} \to Z_\infty \quad a.s.,$$

being $Z_\infty = 1$ if $E(Z_{n,1}) > E(Z_{n,2})$ and $Z_\infty = 0$ if $E(Z_{n,1}) < E(Z_{n,2})$. The proportion of assignments to the best treatment converges to 1. In case of equality between the treatment responses, $E(Z_{n,1}) = E(Z_{n,2})$, Z_∞ has a Beta distribution, $\beta(W_{0,1}, W_{0,2})$.

In [52] a very intuitive bias coin design is presented. The aim is to reach a fixed target allocation, ρ, that depends on the unknown parameters of the design, usually p_1 and p_2.

The target allocation is estimated in each step, and this estimation, $\hat{\rho}_n$, is used as the probability to assign patients. The allocation rule is

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = I\left\{U_{n+1} < \hat{\rho}_n\right\}\,, \tag{1.8}$$

being U_{n+1} a uniform distribution in $[0, 1]$.

With this setting and under the assumptions that $N_{n,1} \to \infty$ and $N_{n,2} \to \infty$, the next strong law is proven,

$$\frac{N_{n,1}}{n} \to \rho\,.$$

A central limit theorem is also obtained,

$$\sqrt{n}\left(\hat{p}_{1,n} - \hat{p}_{2,n}\right) \to N\left(0, \left(\sigma_1 + \sigma_2\right)^2\right)\,,$$

with $\sigma_1 = \sqrt{p_1(1-p_1)}$ and $\sigma_2 = \sqrt{p_2(1-p_2)}$. This design is called, in the sequel, Melfi-Page-Geraldes or simply MPG.

In the MPG design, the allocation rule is clearly dependent on the estimation of the target allocation. A more general family of biased coin designs are the doubly adaptive designs. The idea of a doubly adaptive design (DBCD) was presented and studied in [20] and [21]. The meaning of doubly adaptive is that in each step the proportion of allocations is computed, the desired target allocation is estimated, and the probability of allocation is given by a function depending on these two values. In [20], some conditions to this probability function are given in order to ensure the convergence to the target allocation ρ, and to obtain strong laws and central limit theorems for the treatment effect estimators. Nevertheless, the conditions imposed on the allocation function were too stringent, and in [35] an extension of the DBCD design is proposed, with the same idea, but relaxing some of these conditions. The general performance is quite similar. In each step the allocation proportion is computed and the target is estimated. By means of a function $g : [0, 1] \times [0, 1] \to [0, 1]$ the probability of treatment 1 allocation is given. Imposing some conditions on g, the next results are proven,

$$\frac{N_{n,1}}{n} \to \rho, \quad a.s.$$

$$n^{1/2}(N_{n,1}/n - \rho, \hat{\rho}_n - \rho) \to N(\mathbf{0}, \Sigma), \quad [D],$$

where $\hat{\rho}_n$ is an estimator of the target allocation and Σ is the covariance matrix.

In [35], a family of functions, g_α, is proposed, with the following general expression:

$$g_\alpha(x,y) = \begin{cases} 1, & x = 0, \\ 0, & x = 1, \\ \dfrac{y(y/x)^\alpha}{y(y/x)^\alpha + (1-y)((1-y)/(1-x))^\alpha}, & (x,y) \in (0,\,1) \times [0,\,1], \end{cases} \tag{1.9}$$

with $\alpha \geq 0$.

This function assigns in a deterministic way treatment 1 or 2 if the proportion of allocations to treatment 1 is 0 or 1, respectively. In addition, if $x > y$, then, $g(x,y) < y$ guaranteeing the convergence to the target. The choice of the parameter α has also an important role in the performance of the design. The case $\alpha = 0$ corresponds to the MPG design. The case $\alpha = \infty$ has good asymptotic properties but it is completely deterministic, so it is not recommended. The authors leave the choice of the value of α to the practitioners, but in [62] $\alpha = 2$ is advised.

The Efficient Randomized-Adaptive Design (ERADE), presented in [36], is a biased coin design with the following allocation rule.

$$P(\delta_{n+1} = 1 | \mathcal{F}_n) = \begin{cases} \alpha\hat{\rho}_n, & \text{if } \frac{N_{n,1}}{n} > \hat{\rho}_n; \\ \hat{\rho}_n, & \text{if } \frac{N_{n,1}}{n} = \hat{\rho}_n; \\ 1 - \alpha(1-\hat{\rho}_n), & \text{if } \frac{N_{n,1}}{n} < \hat{\rho}_n. \end{cases} \tag{1.10}$$

The conditional probability of treatment 1 allocation varies depending on the sign of the difference between the current proportion of allocations and the estimation of the target allocation. If the present proportion is equal to the estimation, the probability of assigning treatment 1 is the estimation. If treatment 1 is overallocated, the probability is the estimation penalized by a factor α which is in $[0, 1)$ and if treatment 1 is underallocated,

the probability is larger than the estimation, being also tuned by the parameter α. They recommend selecting the value of the parameter α between 0.4 and 0.7.

The behavior of the allocation rule ensures the convergence of the proportion of treatment 1 allocations, $\frac{N_{n,1}}{n}$, to the target

$$\frac{N_{n,1}}{n} \to \rho, \quad a.s.$$

and a central limit theorem,

$$n^{1/2}(N_{n,1}/n - \rho) \to N(0, \sigma^2), \quad [D].$$

One of the main differences between treatment-adaptive designs and response-adaptive designs is the modification of the target allocation due to the inclusion of the responses in the dynamics of the design. In the first family, all of them are oriented to reach the equality of allocations between both treatments. In response-adaptive designs, the target proportion of treatment allocations varies. In some designs, PTW or DTL for example, the target allocation is given by the design itself, in this sense, they are ad-hoc procedures. In other designs, ERADE or DBCD, for instance, the target allocation, ρ, is chosen by the experimenter at the beginning of the trial, and the design targets this allocation. So, the choice of ρ is fundamental in the properties of the design. In [64] a comparative study of the performance and the optimality of different target allocations is done. A function $f(p_1, p_2)$ is chosen to compare the estimators of the parameters, $\hat{p}_{n,1}$ and $\hat{p}_{n,1}$, and a minimization problem is raised: Find the allocation ratio that minimizes the expected number of failures for a fixed asymptotic variance of $f(\hat{p}_{n,1}, \hat{p}_{n,2})$. Then, the optimal allocation is computed. For instance, in the most usual case, when $f(p_1, p_2) = p_1 - p_2$, the optimal allocation is $\frac{\sqrt{p_1}}{\sqrt{p_1}+\sqrt{p_2}}$, (RSIHR allocation in the sequel). In this paper other limit allocations are also mentioned. On the one hand, Neyman allocation, $\frac{\sqrt{p_1 q_1}}{\sqrt{p_1 q_1}+\sqrt{p_2 q_2}}$, which minimizes the variance of the sample proportions. This allocation also minimizes the sample size when a power level is fixed. On the other hand, the urn allocation, $\frac{q_2}{q_1+q_2}$, which is not optimal with any criterion, but can be interpreted as the relative risk of failure. The most usual

allocation used to compare designs is the RSIHR allocation. It is optimal from the ethical point of view and maintains low levels of variance of the estimators. Neyman allocation is optimal with an inferential criterion but if $p_1 + p_2 > 1$, allocates more patients to the inferior treatment, which is unethical, and the urn allocation, although is not optimal in any sense, is often used for comparisons involving urn designs.

Balance between ethical purposes and good inferential properties is desirable. These two goals are competitive, because in order to improve the ethical aspect of the design, more patients are assigned to the best treatment and then, balance is not possible except for equal performance. This fact decreases the amount of information of the other treatment, usually increasing the variance of the treatment effect estimators which produces lower power levels in the test statistics.

As stated in [33], one of the main issues in response-adaptive randomization is the ability to minimize the expected number of failures without a loss in the power. In line with this, in [32], they derived an explicit relationship between the variance of the proportion of treatment 1 allocations and the power function of the classical test of difference of means. Using a series expansion of the noncentrality parameter, they obtain that the power function is linearly related to $Var(N_{n,1}/n)$, and if the variance increases, the noncentrality parameter decreases, with a loss in the power of the test. This result is strongly linked to the limit allocation, so the variance of $N_{n,1}/n$ could be used as a measure of goodness of the design from the inferential point of view, but only when comparing designs with the same limit allocation.

In [34] a lower bound on the asymptotic variance of the allocation proportion is derived. Using a Cramer-Rao type bound and under asymptotic normality of the allocation proportion, they give a minimum value for the variance. This bound depends on the asymptotic proportion of allocations and it becomes a good tool to compare different designs with the same target allocation. If the proportion of treatment 1 allocations converges almost surely to a value $\rho(p_1, p_2)$ and a central limit theorem can be applied, then the lower bound

for the variance of $N_{n,1}/\sqrt{n}$ is

$$\left(\frac{\delta\rho(p_1,p_2)}{\delta p_1)}\right)^2 \frac{p_1q_1}{\rho(p_1,p_2)} + \left(\frac{\delta\rho(p_1,p_2)}{\delta p_2)}\right)^2 \frac{p_2q_2}{(1-\rho(p_1,p_2))} .$$

For urn allocation, $\frac{q_2}{q_1+q_2}$, the lower bound of the variance is

$$\frac{q_1q_2(p_1+p_2)}{(q_1+q_2)^3} .$$

For Neyman allocation the bound is

$$\frac{1}{4\left(\sqrt{p_1q_1}+\sqrt{p_2q_2}\right)^3}\left(\frac{p_2q_2(q_1-p_1)^2}{\sqrt{p_1q_1}}+\frac{p_1q_1(q_2-p_2)^2}{\sqrt{p_2q_2}}\right)$$

and for RSIHR allocation,

$$\frac{1}{4\left(\sqrt{p_1}+\sqrt{p_2}\right)^3}\left(\frac{p_2q_1}{\sqrt{p_1}}+\frac{p_1q_2}{\sqrt{p_2}}\right) .$$

In DBCD design with $\alpha = 2$, the lower bound is not reached for any of the previous allocations. If $\alpha \to \infty$ the bound is achieved but the design is almost deterministic. However, in ERADE designs, the lower bound is reached for any chosen target allocation. Asymptotic efficiency is the most remarkable property of these designs. In the DTL design, the lower bound for the urn allocation is also reached asymptotically.

The idea of comparing different designs to establish which is the best one is a complex task in response-adaptive randomization, due to the competitiveness between ethical and inferential properties. The usual approximation is looking for a good compromise between both criteria. Either fix one criterion and optimize the other one or a joint optimization of a convex combination of them are the common strategies. In [15], the equivalence between these two approaches is proven.

A deep comparative computational study among response-adaptive designs was done in [62]. The complete randomization design is used as a benchmark to establish the sample size required to have a 90% of power in the classical test for the difference of success probabilities. The comparison is made by using the following two criteria: the total number of failures, as an ethical criterion, and the power of the test, as an inferential

criterion. Firstly, the DTL rule and the PTW rule are compared. As both of them have the same target allocation, the expected number of failures is similar but the DTL rule shows a higher power. So, the superiority of this rule is confirmed. Another comparison is done involving the DBCD design, with three values of the parameter, $\alpha = 0$, $\alpha = 2$ and $\alpha = \infty$. In this comparison the value $\alpha = 2$ seems to be the most advisable, due to the fact that it has very good levels of power and expected failures, with enough level of randomness. Finally, comparing the DTL rule and the DBCD with $\alpha = 2$, the values of both criteria are quite similar, being impossible to determine clearly which is the best one.

In [24] a comparison study is made by using graphical tools. Here, the main idea is to compute two measures, one about ethics and the other one about estimation, drawing the results in a two axis plot. These measures are the expected proportion of allocations in the worst treatment and the root mean-square-error of the estimation of the treatment effects difference. The optimality consists in minimizing at once both measures, so, a global measure of optimality is the distance to the origin. The comparison is made for the most significant response-adaptive designs and they conclude that ERADE designs and DBCD designs have the best performance. They also mention that, among the designs that do not target any optimal allocation, the DTL rule performs the best and is competitive with ERADE and DBCD.

1.2 Randomization based inference

The main objective of a phase III clinical trial is to evaluate the superiority of the treatment under study against the control treatment. In order to make the decision, a statistical test can be used. A frequently used statistical test in this kind of experiments is the classical Wald test for the difference of means,

$$\frac{\hat{p}_{n,1} - \hat{p}_{n,12} - (p_{n,1} - p_{n,2})}{\sqrt{\frac{\hat{p}_{n,1}(1-\hat{p}_{n,1})}{N_{n,1}} + \frac{\hat{p}_{n,2}(1-\hat{p}_{n,2})}{N_{n,2}}}}.$$

Although the distribution of this statistic is unknown in the adaptive case, assuming a population model a normal approximation can be used, in the conditions of [11]. In a population model, patients are assumed to come from two different populations and they are chosen from these populations by a random sampling procedure. However, a peculiarity of clinical trials is the sampling procedure, that differs from the usual one in the majority of statistical experiments. The population of reference is the people who meet suitable requirements in order to participate in the trial. In addition, they must be localized in the area of the centers where the trial is carried on. After that, during the different phases of the study, patients could decline participating in the trial or clinicians could have a negative opinion about a patient and they can rule him out. So, in fact, the sample is formed by all the patients who meet eligibility criteria, are localized, give their consent and are accepted by the clinicians. After that, these patients are allocated in the treatments by the randomized allocation rule of the design and the study is carried out. Note that the only randomness of the whole process is due to the random allocation rule and the sampling process is quite deterministic. This is the reason why the acceptance of a population model, with two different populations and random sampling of patients, is controversial.

An alternative to the population model is the randomization model, which uses the randomization as a basis for inference. A good explanation about randomization based inference (RBI) is given in [18]. Given the experimental data, x, for which no parametric assumption is done, a statistic $T(x)$ is computed. Then, the data are permuted repeatedly in a manner consistent with the random assignment procedure, and the statistic is computed for each of the resulting data permutations. Suppose that a one-sided significance test with significance level α is applied to accept or reject a null hypothesis. The null hypothesis for a randomization test is that there is no difference in the observed outcomes whatever the allocation, which can be seen as a nonparametric counterpart of the parametric hypothesis of equality of success probabilities. RBI proceeds by computing the

probability that the random assignment procedure provides a reallocation for which the test statistic is greater than the observed value $T(x)$. In other words, the p-value associated with $T(x)$ is computed. If this probability is smaller than α, the null hypothesis is rejected.

The origin of randomization in the design of experiments and the concept of randomization test can be found in Fisher's work in the 1930's. Since then, it has been a controversial statistical issue. In [8] a good review about the origin, development and criticisms of randomization is done. Focusing more on permutation tests, in [56], along with a solid theoretical basis for permutations tests, some classical criticisms about this inferential procedure are presented, but their adequacy in some cases is vindicated.

In the context of clinical trials, following [63], the null hypothesis of a randomization test is no difference between both treatments. So, the assignment of the treatments does not affect the responses of the patients, and they are assumed fixed. Then, the only randomness left comes from the patients' allocations.

Another important statistical issue when randomization tests are used concerns the use of conditional or unconditional tests. Conditional tests assume that the decision must be made with the p-value of the test statistic conditioned to the number of patients allocated in treatment 1, that is, the distribution of $T(x)|\{N_{n,1} = n_1\}$ must be calculated instead of the unconditional distribution of $T(x)$. The use of conditional tests is advised by Cox in [16] when the number of patients allocated to each treatment is ancillary, as it happens for treatment-adaptive designs. However, with response-adaptive designs, the distribution of $N_{n,1}$ depends on the success probabilities and, so, it is not ancillary and the use of conditional tests is not justified with Cox's arguments. But in [9], where results in [74] are analyzed, the use of conditional tests is considered more appropriate also for response-adaptive designs, but always under heuristic arguments.

RBI requires the computation of the test statistic for all the possible allocations of patients and, also, the calculation of the probability for each possible arrangement of al-

locations, when arrangements are not equally probable. This could be computationally unfeasible, and becomes the main drawback of RBI. In order to avoid this problem, several techniques have been presented. In [49] and [57], Monte Carlo methods are used to compute p-values via approximation. Another usual technique consists in programming algorithms to shorten the calculations of all the permutations, using recurrence relations. In [31] an algorithm to compute the exact distribution of the randomization test statistic for the Smith class of designs is given and in [50] another algorithm is presented based on networks in order to obtain the exact distribution for Wei's urn design. It should be mentioned that these recursive algorithms are very tied to the particular class of designs analyzed, because the recursive rules are based on relations originated from the allocation probabilities. For response-adaptive designs these relations could be more complicated, due to the inclusion of responses in the allocation rule. Less work has been done in this field. In [74] the exact distribution of the test is computed for the play-the-winner rule and in [23] for the drop-the-loser rule. In some cases, with very large sample sizes, this recurrence relations are computationally unfeasible, so, an alternative is to obtain the asymptotic distribution of the test statistic. As the responses are a deterministic set, ensuring an asymptotic distribution for all the possible score sets is almost impossible and some conditions on them are required. In addition, all the procedures for obtaining the asymptotic distributions of these statistics depend strongly on the randomization rule. For instance, in [71] a central limit theorem is obtained for the randomization statistic when Wei's urn design is applied. They also prove with a counterexample that for some set of scores the statistic is not normal if Efron's design has been applied.

1.3 Covariates in clinical trials

In the design and analysis of a clinical trial, there is additional patient information that should be taken into account. The covariates or prognostic factors give us more information about patients, and can influence the patients' response to the treatment. Gender, age,

clinical centre or some health indicators related to the disease could be associated to the response that we are studying. An imbalance between covariate levels could produce a bias in the study. Randomization itself protects against these imbalances, because the random rule asymptotically distributes equally patients between treatments. For instance, in [63] a measure of covariate imbalance is presented,

$$\Delta F_n := \frac{\sum_{i=1}^{n} \delta_i H_i}{N_{n,1}} - \frac{\sum_{i=1}^{n} (1-\delta_i) H_i}{N_{n,2}}, \tag{1.11}$$

and they give conditions to prove $P(|\Delta F_n| > \varepsilon) \to 0$. But this is an asymptotic property and in moderate sample sizes large imbalances could appear, biasing the study. Several solutions have been presented in order to avoid or mitigate this bias, as including the covariates in the design, or analyzing the results taking into account these covariates. Good reviews of the use of covariates in adaptive designs can be found in Chapter 4 of [63] and in [65].

Stratification is a common approach to include the covariate information. In a stratified design, the strata are defined using the different covariate levels. These strata are considered separately, each one with its proper randomization rule. When a patient arrives to the study, he is assigned to a stratum according to his covariate information and the rule of this stratum is used to assign him to a treatment. Stratification also could be done in a post-randomization stage, comparing the different treatments within each strata and then combining the whole information in a unique procedure.

Another different approach is the so-called covariate-adaptive randomization. In this setting, the covariate information is included in the general randomization rule, to minimize the imbalance between the different groups. Formally, the $(n+1)$-th patient is allocated depending on the previous allocations and covariates and his/her covariates, so $\mathcal{F}_n = \sigma(\mathcal{G}_n, \mathcal{H}_{n+1})$.

The use of covariates in randomization rules was first proposed in [59]. The performance of the rule is very similar to the Efron's biased coin desing, but measuring covariate

imbalances instead of treatment imbalances. The main idea is to define an imbalance metric, D_n, which measures the total imbalance between groups, with a weighted sum of the imbalance in each group. Then, the allocation rule is

$$P(\delta_{n+1}=1|\mathcal{F}_n)=\begin{cases} p & if\ D_n<0\,,\\ 1-p & if\ D_n>0\,,\\ \frac{1}{2} & if\ D_n=0\,. \end{cases}$$

In [37] a version of the Pocock and Simon's model is presented an its properties are studied. The lack of theoretical work in covariate-adaptive randomization has been remarked in several papers, for instance, in [37] and in [65].

The model-based approach for handling covariates was proposed in [10] and in [2]. In these papers, a linear model is assumed for the responses of the patients. This model considers the linear influence of the allocations and the covariates of interest. Then, if n patients have been allocated, the $(n+1)$-th patient is allocated to minimize the variance of the treatment effect. In [10], this version of the design was presented, being a deterministic rule, and in [2] a randomized version was studied. In [70] the theoretical basis for these designs was developed.

When the $(n+1)$-th patient is allocated depending on the previous responses, allocations, covariates and, also, her/his own covariates, the information is collected as $\mathcal{F}_n=\sigma(\mathcal{Z}_n,\,\mathcal{G}_n,\,\mathcal{H}_{n+1})$. Then, we say that the rule is covariate adjusted response-adaptive (CARA). A seminal paper which relates covariates and adaptive designs is [66], where a treatment effect is used to allocate the next patient to a treatment. More precisely, they consider two treatments with dichotomous responses, and they assume a logistic response model with parameters α, β, $\boldsymbol{\gamma}=(\gamma_1,\,\ldots,\,\gamma_M)^t$, $\boldsymbol{\theta}=(\theta_1,\,\ldots,\,\theta_M)^t$:

$$logit(P(Y_n=1))=\alpha+\beta\delta_n+\mathbf{H}_n^t\boldsymbol{\gamma}+\delta_n\mathbf{H}_n^t\boldsymbol{\theta}\,,$$

where $\mathbf{H}_n$ is the vector which contains the patient's covariate information. Given the

maximum-likelihood estimators of the parameters, they allocate the next patient using

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = \frac{1}{1 + e^{-(\hat{\beta}_n + \boldsymbol{H}_{n+1}^t \hat{\boldsymbol{\theta}}_n)}}$$

which is a function of the estimated covariate-adjusted odds ratio between treatments. They show, using a simulation study, that this rule skews allocation to the best treatment and has less failures than equal allocation.

Several asymptotic properties of a class of covariate adjusted response-adaptive designs have been studied in [78]. These designs assume a parametric model and have a target allocation, which is a function of some parameters of the model. The patient's response to treatment j, Y_{nj}, satisfies

$$E[Y_{n,j}|\mathcal{H}_n] = p(\theta_j, \mathbf{H}_n)\,.$$

where θ_j are the parameters of the model. Then, the allocation rule is a function of the covariates, which are known, and of the parameters, which can be estimated using the accrued information.

$$P(\delta_{n+1} = 1|\mathcal{F}_n) = \pi(\hat{\theta}, \mathbf{H}_n)$$

Asymptotic results for estimators and for the proportion of successes (by stratum, by treatment) are obtained for this class of designs, which includes the setting in [66].

Chapter 2

Klein urn design

A randomized controlled clinical trial is a statistical experiment to compare the efficacy of a new treatment, say treatment 1, with respect to a control treatment, say treatment 2. The control treatment is the best clinical practice known or a placebo. Patients are assumed to arrive sequentially to the experiment and, following a random rule, are allocated in treatment 1 or treatment 2. When the allocation rule uses previous information to allocate the following patient, the design is called adaptive. Adaptive designs could be classified according to the previous information which is used. Following [63], we say that a design is treatment-adaptive if it uses information about past allocations, and we say that it is response-adaptive if it also uses information of the past responses of patients to treatments. The Ehrenfest urn design, see (1.4), is a treatment-adaptive design that uses the previous allocations in order to assign treatment 1 or 2 to the next patient. Its limit allocation for each treatment is 1/2. If the result of a treatment can be observed before the arrival of the next patient, the information accrued from the previous individuals can be used to improve the performance of the experiment. Some modifications of the Ehrenfest urn design which use this information are presented in this chapter. In the first scenario, S1, a treatment is reinforced when is successful. This rule mimics the Randomly Reinforced Urn model, RRU, see (1.7). In the second scenario, S2, a treatment is reinforced if it is a success or if the other treatment is a failure. This is the randomized Play-The-Winner rule,

PTW, which has been profusely studied, see (1.5). In scenario S3 a treatment is reinforced if the other treatment is applied and it is a failure. This rule is similar to the Drop-The-Loser rule, DTL, see (1.6). Exact and asymptotic properties of these three scenarios are obtained, and their performance is compared with their inspiring rules. Scenario S3 has a similar behavior to the DTL rule in the number of failures and power and also in the variability of allocations. Besides, it has the same target allocation as the DTL rule, which is widely accepted as a competitive response-adaptive design, see [34] and [24]. Therefore, scenario S3 deserves a thorough study, which is done in the last section of this chapter. The urn composition of S3 is a homogeneous Markov chain which was studied by Klein in [41] out of the context of clinical trials, as a generalization of the Ehrenfest urn model. This scenario S3 will be called Klein urn design in what follows.

The contents of this chapter have been published in [27] and [26].

2.1 Response-adaptive designs based on Ehrenfest urn

The Ehrenfest urn design, see, for instance, (1.4) or [12] for the original paper, is a treatment-adaptive design that uses the previous information in the following way. Initially the urn contains w balls of each type, that is, $(W_{0,1}, W_{0,2}) = (w, w)$. When a new patient arrives, a ball is drawn from the urn, the patient is allocated to the treatment associated with its type and a ball of the other type is added to the urn. Observe that the total number of balls remains fixed after each replacement. So that, for each n, $W_{n,1} + W_{n,2} = 2w$ and $W_{n,1}$ describes completely the composition of the urn. The composition of the urn depends only on the number of times that each treatment has been applied.

Mean and variance can be calculated, at any stage n of the procedure, for the processes $\{W_{n,1}\}$ and $\{N_{n,1}\}$ associated with the Ehrenfest urn design, see [5], and are

$$E[W_{n,1}] = w, \qquad Var[W_{n,1}] = \frac{w}{2}\left(1 - \left(1 - \frac{2}{w}\right)^n\right),$$

and also,

$$E\left[\frac{N_{n,1}}{n}\right]=\frac{1}{2},\qquad Var\left[\frac{N_{n,1}}{n}\right]=\frac{w}{8n^2}\left(1-\left(1-\frac{2}{w}\right)^n\right).$$

Observe that the process $\{W_{n,1}\}$ is a time homogeneous Markov Chain with state space $E=\{0,\,1,\,\ldots,\,2w\}$ and transition matrix $P=(p_{i,j})$, where

$$p_{i,j}=\begin{cases}1-\frac{i}{2w}, & j=i+1;\\ \frac{i}{2w}, & j=i-1;\end{cases}\qquad i=0,\,1,\,\ldots,\,2w. \tag{2.1}$$

This property can be exploited, see [6], to obtain strong laws and central limit theorems. In particular, the following central limit theorem holds,

$$\sqrt{n}\left(\frac{N_{n,1}}{n}-\frac{1}{2}\right)\to N\left(0,\frac{\sigma^2}{(2w)^2}\right),\quad [D]$$

where σ^2 can be expressed in terms of the eigenvalues and eigenvectors of the transition matrix of the Markov Chain $\{W_{n,1}\}$.

Three modifications of the Ehrenfest urn design are introduced in this section, in which depending on the response of the patient, the extracted ball is replaced or not. In the three scenarios we consider an urn that, initially, contains $W_{0,1}$ balls of type 1, and $2w-W_{0,1}$ balls of type 2. $W_{0,1}$ could be any random variable with values in $E=\{0,\,1,\,\ldots,\,2w\}$, but usually we assume that $W_{0,1}=w$. The total number of balls remains constant, $W_{n,1}+W_{n,2}=2w$, for any n, so as in the Ehrenfest urn model, $W_{n,1}$ describes completely the state of the urn. $\{W_{n,1}\}$ is a discrete time stochastic process which takes values in the set $E=\{0,\,1,\,\ldots,\,2w\}$. As the urn replacement policy only depends on the past history of the process via the last state, $\{W_{n,1}\}$ is clearly a Markov chain. We present a description of each scenario along with the transition probabilities of the Markov chain $\{W_{n,1}\}$.

2.1.1 Scenario 1 (S1 design)

In scenario $S1$ a treatment is reinforced when it is successful. As stated in the introduction, the RRU designs, (1.7), uses an unbounded urn, and the proportion of patients allocated in

the best treatment converges to 1 almost surely. When both treatments perform equally, it converges to a beta distribution $\beta(w,\, w)$.

The reinforcement policy has to be adapted to obtain an urn with a constant number of balls, $2w$. When the urn is in an interior state, that is, when $W_{n,1} \in E \setminus \{0,\ 2w\}$, if the treatment applied is a success, we add a ball of its type and we remove a ball of the other type. If the treatment is a failure, the composition of the urn remains unchanged. The transition matrix of $\{W_{n,1}\}$ derived from the S1 rule is $P = (P_{i,j})\ i, j \in E$, where

$$p_{i,j} = \begin{cases} p_1 \frac{i}{2w}, & j = i+1; \\ q_1 \frac{i}{2w} + q_2(1 - \frac{i}{2w}), & j = i; \\ p_2(1 - \frac{i}{2w}), & j = i-1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 1, \ldots, 2w-1\,. \tag{2.2}$$

When $W_{n,1} = 0$ or $W_{n,1} = 2w$, additional rules are needed. If the treatment applied is a success, the urn remains unchanged. If it is a failure, we remove a ball of its type and we add a ball of the other type. So, the states 0 and $2w$ are considered as semi-reflecting barriers, that is

$$\begin{aligned} p_{0,0} &= p_2, \qquad p_{0,1} = q_2 \\ p_{2w,2w} &= p_1, \quad p_{2w,2w-1} = q_1. \end{aligned} \tag{2.3}$$

2.1.2 Scenario 2 (S2 design)

In scenario $S2$ a treatment is reinforced if it is a success or if the other treatment is a failure. This rule corresponds with the randomized PTW rule, (1.5), presented in the introduction. In the PTW rule an unbounded number of balls in the urn is also considered and the proportion of patients allocated in treatment 1 converges to $q_2/(q_1 + q_2)$; that is, the ratio of allocations to a treatment converges to its relative risk of failure.

The reinforcement policy has to be adapted to obtain an urn with a fixed number of balls, $2w$. When the urn is in an interior state, $W_{n,1} \in E \setminus \{0,\ 2w\}$, we add a ball of type 1 and remove a ball of type 2 if treatment 1 is applied and it is a success or if treatment 2 is applied and it is a failure; treatment 2 is reinforced in a similar way: if treatment 2 is

applied and it is a success or if treatment 1 is applied and it is a failure. In this case, the transition matrix P of the MC $\{W_{n,1}\}$ is

$$p_{i,j} = \begin{cases} p_1 \frac{i}{2w} + q_2(1 - \frac{i}{2w}), & j = i+1; \\ p_2(1 - \frac{i}{2w}) + q_1 \frac{i}{2w}, & j = i-1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 1, \ldots, 2w-1\,. \tag{2.4}$$

When $W_{n,1} = 0$ or $W_{n,1} = 2w$, we proceed as in (2.3):

$$p_{0,0} = p_2, \qquad p_{0,1} = q_2$$
$$p_{2w,2w} = p_1, \quad p_{2w,2w-1} = q_1.$$

2.1.3 Scenario 3 (S3 design)

In scenario $S3$ a treatment is reinforced if the other treatment is applied and it is a failure. This rule is similar to the DTL rule, (1.6). The DTL rule has the same allocation limit as the PTW rule; that is, the proportion of patients allocated to a treatment converges to its relative risk of failure.

The transition matrix P under the S3 design is

$$p_{i,j} = \begin{cases} q_2(1 - \frac{i}{2w}), & j = i+1; \\ p_1 \frac{i}{2w} + p_2(1 - \frac{i}{2w}), & j = i; \\ q_1 \frac{i}{2w}, & j = i-1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 0, \ldots, 2w. \tag{2.5}$$

The urn remains unchanged if the treatment applied is a success. If it is a failure, we remove a ball of this type and we add a ball of the other type. Note that, in this scenario, the barrier conditions of the previous scenarios implicitly hold, due to the fact that the general rule could be applied in the border states. Transition matrix (2.5) was already considered in [41].

Remark 2.1.1. The semi-reflecting barrier conditions established in S1, S2 and S3 designs seem quite logical in the spirit of a clinical trial. If $W_{n,1} = 0$ the urn contains $2w$ balls of

type 2, and treatment 2 is applied until a failure happens. If $W_{n,1} = 2w$ the urn contains $2w$ balls of type 1, and treatment 1 is applied until a failure happens. On the other hand, absorbent barriers, the case that the Markov chain gets stuck when it hits a barrier, would force to apply the same treatment once the barrier is reached until the end of the trial, which is a deterministic rule for a clinical trial. This introduces biases in the study, so it is not a viable alternative. Completely reflecting barriers, the case that the chain automatically moves to the adjacent state when it hits a barrier, makes the urn process to be deterministic when the barrier is reached. In these cases, regardless the response of the patient, the urn would move like if a failure had happened, losing the connection between the urn and the performance of the treatments.

As stated in [32], "Any new procedures proposed should fully investigate operating characteristics, including the target allocation, expected failure rate, variability, and power". Also in [32], the study of asymptotic properties is enhanced. After presenting the three designs, checking their properties should be an mandatory step, comparing them with another designs in the literature. Comparative studies rely heavily on asymptotical properties or on simulation studies. Checking asymptotic properties of the designs is essential, as mentioned in [32], "is important to check the accuracy of the asymptotic approximations when using these theoretical results to compare designs". So, it is of interest and necessary to carry out a study of the asymptotic properties of S1, S2 and S3 designs.

On the other hand, it is also important to know some exact information about the operating characteristics in any step n of the process. Some clinical trials could have small sample size, which prevents the use of asymptotic results. Besides this, even in trials involving large quantities of patients, information about the behavior in early steps is useful. But, in general, exact values, for example the mean and the variance of the number of allocations for each n, are difficult to obtain when adaptive designs are applied, due to the complicated correlation structure generated between allocations and observed responses. However, in S1, S2 and S3 designs, some exacts results are obtained.

The next proposition deals with the asymptotic behavior of the $\{W_{n,1}\}$ process, which is based on some classical results of Markov chains.

Proposition 2.1.1. *Under the three scenarios, the process* $\{W_{n,1}\}$ *has the following properties.*

a) It is an aperiodic and irreducible Markov chain and there exists a stationary distribution $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$

b) A strong law holds

$$\frac{1}{n}\sum_{k=0}^{n-1} W_{k,1} \to \boldsymbol{\pi}^*, \quad a.s.$$

where $\boldsymbol{\pi}^* := \sum_{i\in E} i\pi_i$ *is the mean value of the stationary distribution. Besides, for* $m > 1$,

$$\frac{1}{n}\sum_{k=1}^{n} W_{k,1}^m \to \boldsymbol{\pi}_m^*, \quad a.s.$$

where $\boldsymbol{\pi}_m^* := \sum_{i\in E} i^m \pi_i$.

c) A central limit theorem holds

$$\frac{1}{\sqrt{n}}\left(\sum_{k=0}^{n-1} W_{k,1} - n\boldsymbol{\pi}^*\right) \to N(0,\ \sigma^2).$$

where $\sigma^2 = \lim_{n\to\infty} \frac{1}{n} Var\left[\sum_{k=1}^{n} W_{k,1}\right]$.

Proof. The chain is clearly irreducible and aperiodic under the three scenarios. Since the state space is finite, the chain is positive recurrent. The existence of the stationary distribution is given by the Lemma III.2.1 in [38] and will be denoted $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$. For the strong law, note that

$$\frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\sum_{i\in E} i\, n(X_k = i)$$

where $n(X_k = i)$ is the number of visits to the state i. Following proposition 2.21 in [14], the result is immediate. Finally, the c) part is a particular case of the so called CLT for Markov chains; see, for instante Theorem 10.2 in [17]. □

In the following proposition we give an explicit expression of the distribution probabilities.

Proposition 2.1.2. *For the Markov chain $\{W_{n,1}\}$, the stationary distribution $\boldsymbol{\pi}$ satisfies:*

a) for scenario 1, (S1),

$$\begin{aligned}\frac{\pi_i}{\pi_0} &= \frac{q_2}{p_2}\left(\frac{p_1}{p_2}\right)^{i-1}\frac{2w}{\binom{2w-1}{i-1}(2w-i)},\ i=1,\dots,2w-1\\ \frac{\pi_{2w}}{\pi_0} &= \frac{q_2}{q_1}\left(\frac{p_1}{p_2}\right)^{2w-1}.\end{aligned}$$

b) For scenario 2, (S2),

$$\frac{\pi_i}{\pi_0} = \frac{\prod_{j=0}^{i-1} q_2+(p_1-q_2)\frac{j}{2w}}{\prod_{j=2w-i}^{2w-1} q_1+(p_1-q_2)\frac{j}{2w}},\ i=1,\dots,2w\,.$$

c) For scenario 3, (S3), $\boldsymbol{\pi}$ is the probability mass distribution of a binomial distribution with parameters $2w$ and $q_2/(q_1+q_2)$.

Proof. As the transition matrix for the S1, S2 and S3 is tridiagonal, following the example (b) of page 396 in [22], we obtain,

$$\frac{\pi_i}{\pi_0}=\prod_{j=0}^{i-1}\frac{p_{j,j+1}}{p_{j+1,j}},\ i=1,\dots,2w,\ \pi_0=\frac{1}{1+\sum_{i=1}^{2w}\frac{\pi_i}{\pi_0}}\quad.$$

If we substitute in this expression (2.2), (2.4) and (2.5), the transition probabilities of each scenario, we prove the three results. In S1, for each state $i<2w$, we can use $p_{j,j+1}=p_1\frac{j}{2w}$ and $p_{j+1,j}=p_2\left(1-\frac{j+1}{2w}\right)$ so plugging these expressions in the general equation, and noting that $p_{0,1}=q_2$ due to the barrier conditions, we have that

$$\begin{aligned}\prod_{j=0}^{i-1}\frac{p_{j,j+1}}{p_{j+1,j}}=\frac{q_2\prod_{j=1}^{i-1}p1\frac{j}{2w}}{\prod_{j=0}^{i-1}p_2\left(1-\frac{j+1}{2w}\right)} &= \frac{q_2}{p_2}\left(\frac{p_1}{p_2}\right)^{i-1}\frac{(i-1)!}{\prod_{j=0}^{i-1}(2w-(j-1))}\\ &= \frac{q_2}{p_2}\left(\frac{p_1}{p_2}\right)^{i-1}\frac{2w}{\binom{2w-1}{i-1}(2w-i)}\,.\end{aligned}$$

The closed expression for $i = 2w$, S2 and S3 are obtained with the same procedure. □

Remark 2.1.2. Observing the expressions in the previous proposition, we note that when $p_1 = p_2 = p$, the stationary distribution is symmetric for the three scenarios. So that, $\boldsymbol{\pi}^* = w$. Besides this, if the starting distribution is symmetric, then the probability distributions of $\{W_{n,1}\}$ are symmetric for any $n \geq 0$, due to the symmetry of the transition matrix for each scenario. For scenario S2, when $p = 0.5$ we have that $\boldsymbol{\pi}$ follows the uniform distribution on the set $\{0, 1, \ldots, 2w\}$.

The expressions for the stationary distribution in the last proposition allows us to implement them and obtain a visual approach about how the probability mass is distributed. In Figures 2.1 and 2.2, the stationary distribution for S1 and S2 is shown for selected values of p_1 and p_2.

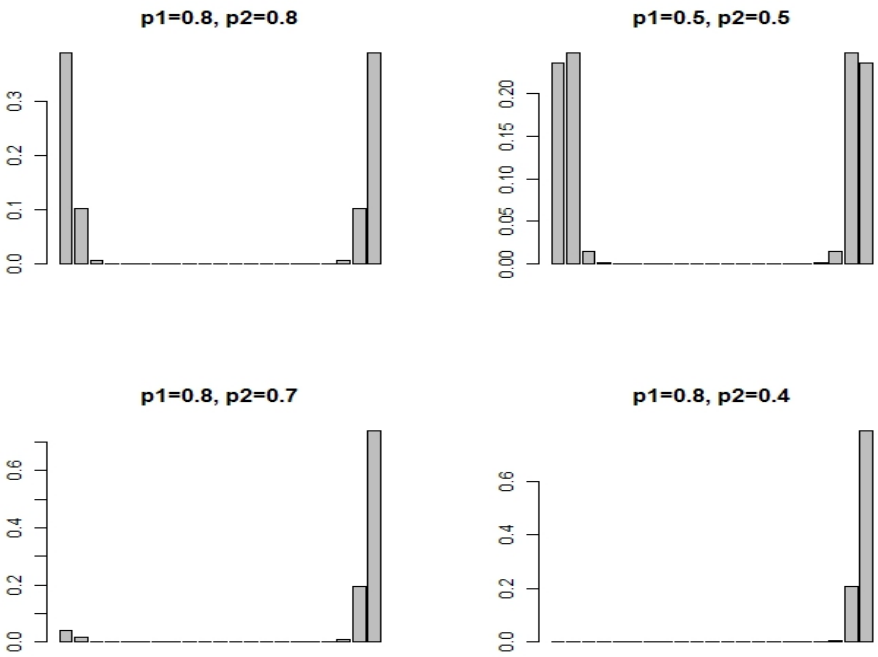

Figure 2.1: Stationary distributions of $\{W_{n,1}\}$ in S1

In S1 the different stationary distributions can be divided into two different groups,

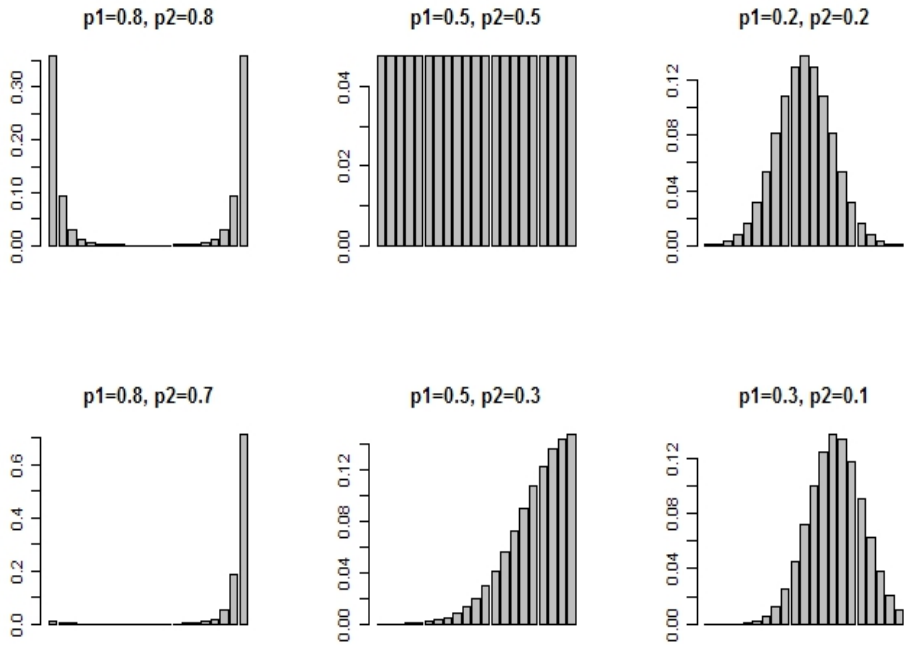

Figure 2.2: Stationary distributions in S2

being Figure 2.1 a sample of them. In the case of different success probabilities, the probability mass is completely accumulated in larger states of E, that is, a larger amount of balls of the best treatment is more probable in the stationary case. If the difference between both probabilities is small, then a small amount of probability appears in the border state of the worst treatment, as we can see in the 0.8-0.7 case shown in Figure 2.1. When both are equal, the distribution is symmetric, as we know from the explicit expressions, and completely accumulated in the smaller and larger states of the chain.

For the case S2, we can observe more variety of distributions. Under equal success probabilities, if $p_1 = p_2 > 0.5$ then the distribution has a U form like in the previous scenario. If $p_1 = p_2 < 0.5$ the distribution has a bell shape, concentrating more probability in the central states, due to the low success probabilities. The limit case $p_1 = p_2 = 0.5$ is the uniform distribution, as stated in Remark 2.1.2. In the case of different success probabilities, the distribution is skewed towards the best treatment, being this more pronounced

with larger success probabilities.

A closed expression for the mean and variance of the stationary distribution for each scenario is of interest. In the following proposition, by means of a recurrence relationship, explicit expressions of $\boldsymbol{\pi}^*$, the mean of the stationary distribution is presented for each scenario.

Proposition 2.1.3. *We consider the Markov chain* $\{W_{n,1}\}$ *and assume that* $W_{0,1} = w$. *Then,*

a) for scenario 1, S1,

$$\boldsymbol{\pi}^* = \frac{2w}{p_1 + p_2}(p_2 + \pi_{2w} - \pi_0)$$

b) for scenario 2, S2,

$$\boldsymbol{\pi}^* = \frac{w}{q_1 - p_2}(q_2 - p_2 + p_2\pi_0 - p_1\pi_{2w})$$

c) for scenario 3, S3,

$$\boldsymbol{\pi}^* = 2w\frac{q_2}{q_1 + q_2}$$

Proof. Once the response of the kth patient is obtained, we denote as I_k^+ the indicator variable of adding one ball of type 1 to the urn and I_k^- the indicator variable of removing one ball of type 1 from the urn. We can write the following recurrence equation

$$W_{n+1,1} = W_{n,1} + I_{n+1}^+ - I_{n+1}^-. \tag{2.6}$$

Note that

$$\begin{aligned} E[I_{n+1}^+|W_{n,1}] &= \sum_{i=1}^{2w-1} p_{i,i+1} I_{\{W_{n,1}=i\}} + q_2 I_{\{W_{n,1}=0\}}, \\ E[I_{n+1}^-|W_{n,1}] &= \sum_{i=1}^{2w-1} p_{i,i-1} I_{\{W_{n,1}=i\}} + q_1 I_{\{W_{n,1}=2w\}}. \end{aligned}$$

Expectations are taken in (2.6) and we have that

$$E[W_{n+1,1}] = E[W_{n,1}] + \sum_{i=1}^{2w-1} (p_{i,i+1} - p_{i,i-1}) p_{w,i}^n + q_2 p_{w,0}^n - q_1 p_{w,2w}^n. \tag{2.7}$$

On the other hand, for $i = 1, \ldots, 2w-1$, and depending on the scenario,

$$p_{i,i+1} - p_{i,i-1} = \begin{cases} -p_2 + \dfrac{p_1+p_2}{2w} i, & \text{for } S1\ , \\ -(p_2 - q_2) + \dfrac{p_1 - q_2}{w} i, & \text{for } S2, \\ q_2 - \dfrac{q_1+q_2}{2w} i, & \text{for } S3\ . \end{cases} \tag{2.8}$$

So that, $p_{i,i+1} - p_{i,i-1} = a_1 + b_1 i$, where a_1 and b_1 are constants determined in (2.8) depending on the scenario. Then, (2.7) becomes:

$$E[W_{n+1,1}] = (1+b_1)E[W_{n,1}] + a_1 + (q_2 - a_1)p^n_{w,0} - (q_1 + a_1 + 2wb_1)p^n_{w,2w}, \tag{2.9}$$

which is the recurrence relation for $E[W_{n,1}]$ that we were looking for.

Taking limits in (2.9) we have

$$\boldsymbol{\pi}^* = \frac{a_1 + (q_2 - a_1)\pi_0 - (q_1 + a_1 + 2wb_1)\pi_{2w}}{-b_1} \tag{2.10}$$

and a), b) and c) follow plugging in (2.10) the coefficients given in (2.8).

□

Remark 2.1.3. Using the same procedure we can get also a recurrence expression for $E[W^2_{n+1,1}]$. From (2.6) we have

$$W^2_{n+1,1} = W^2_{n,1} + I^+_{n+1} + I^-_{n+1} + 2W_{n,1}(I^+_{n+1} - I^-_{n+1}), \tag{2.11}$$

Taking expectations in (2.11) we obtain

$$\begin{aligned} E_w[W^2_{n+1,1}] \;=\; & E_w[W^2_{n,1}] + \sum_{i=1}^{2w-1} (p_{i,i+1} + p_{i,i-1})p^n_{w,i} + q_2 p^n_{w,0} + q_1(1-4w)p^n_{w,2w} \\ & +2\sum_{i=1}^{2w-1} i(p_{i,i+1} - p_{i,i-1})p^n_{w,i} \quad . \end{aligned} \tag{2.12}$$

Note that, for $i = 1, \ldots, 2w-1$, and depending on the scenario,

$$p_{i,i+1} + p_{i,i-1} = \begin{cases} p_2 + \dfrac{p_1 - p_2}{2w} i, & \text{for } S1, \\ 1, & \text{for } S2, \\ q_2 + \dfrac{q_1 - q_2}{2w} i, & \text{for } S3. \end{cases} \tag{2.13}$$

So that, $p_{i,i+1} + p_{i,i-1} = a_2 + b_2 i$, where a_2 and b_2 are constants determined in (2.13) for each scenario. Now we can rewrite (2.12) as follows,

$$\begin{aligned} E[W^2_{n+1,1}] &= (1+2b_1)E[W^2_{n,1}] + a_2 + (b_2 + 2a_1)E[W_{n,1}] \\ &\quad +(q_2 - a_2)p^n_{w,0} \\ &\quad +(q_1 - a_2 - 2b_2 w - 4a_1 w - 8b_1 w^2 - 4wq_1)p^n_{w,2w} \end{aligned} \tag{2.14}$$

which is the recurrence relation for $E[W^2_{n,1}]$.

Remark 2.1.4. A closed expression for the solution of (2.9) is easy to obtain for the S3 design:

$$E[W_{n,1}] = \boldsymbol{\pi}^* + r^n(w - \boldsymbol{\pi}^*),$$

where $r = (1 - \dfrac{q_1 + q_2}{2w})$. As $N_{n,1} = \displaystyle\sum_{k=1}^{n} \delta_k$, we have

$$E[N_{n,1}] = E\left[\sum_{k=1}^{n} \delta_k\right] = \sum_{k=1}^{n} E[E[\delta_k|\mathcal{F}_{k-1}] = \frac{1}{2w}\sum_{k=0}^{n-1} E[W_{k,1}] \tag{2.15}$$

and then, plugging the previous expression for $E[W_{k,1}]$ in 2.15, we obtain

$$E[N_{n,1}] = n\frac{\boldsymbol{\pi}^*}{2w} + \frac{1-r^n}{1-r}\left(\frac{1}{2} - \frac{\boldsymbol{\pi}^*}{2w}\right).$$

Proposition 2.1.4. *We have the next strong laws,*

a) for scenario 1, S1

$$\frac{N_{n,1}}{n} \to \frac{(p_2 + \pi_{2w} - \pi_0)}{p_1 + p_2} \quad a.s.$$

b) for scenario 2, S2

$$\frac{N_{n,1}}{n} \to \frac{(q_2 - p_2 + p_2\pi_0 - p_1\pi_{2w})}{2(q_1 - p_2)} \quad a.s.$$

c) for scenario 3, S3

$$\frac{N_{n,1}}{n} \to \frac{q_2}{q_1 + q_2} \quad a.s.$$

Proof. From Proposition 2.1.1, we have the almost sure convergence of the $\{W_{n,1}\}$ process. Applying Theorem 1 in [6], we have the following result for the three scenarios.

$$\frac{N_{n,1}}{n} \to \frac{\boldsymbol{\pi}^*}{2w}, \quad a.s.$$

Using Proposition 2.1.3, we have the explicit values of $\boldsymbol{\pi}^*$, and we get the result. □

Remark 2.1.5. Observing the limiting values of the proportion of values, some facts could be observed. On the one hand, in S1 and S2 designs, this limit depends strongly on the stationary probabilities of the border states of the urn, and this dependency disappears in the third scenario. On the other hand, the limit in S3 is a well known value, called urn allocation and is the limiting allocation of some designs like DTL and PTW. As stated in the introduction, this value represents the relative risk of failure.

After presenting and studying the properties of the three scenarios, we compare them with their inspiring rules, to check their goodness. We have done a simulation study comparing scenarios S1, S2 and S3 with the RRU, PTW and DTL rules. Based on the previous theory, we know the almost sure convergence of the process of treatment 1 assignments, $\{N_{n,1}\}$. We have simulated for each design 1000 replications of a trial with 200 patients. We have plotted the number of allocations in treatment 1, to see the performance of the design, and also we have checked the convergence to get an idea of the variance of this process under the different rules. The results vary with different success probabilities, we have chosen some pairs of probabilities that summarize the overall behavior.

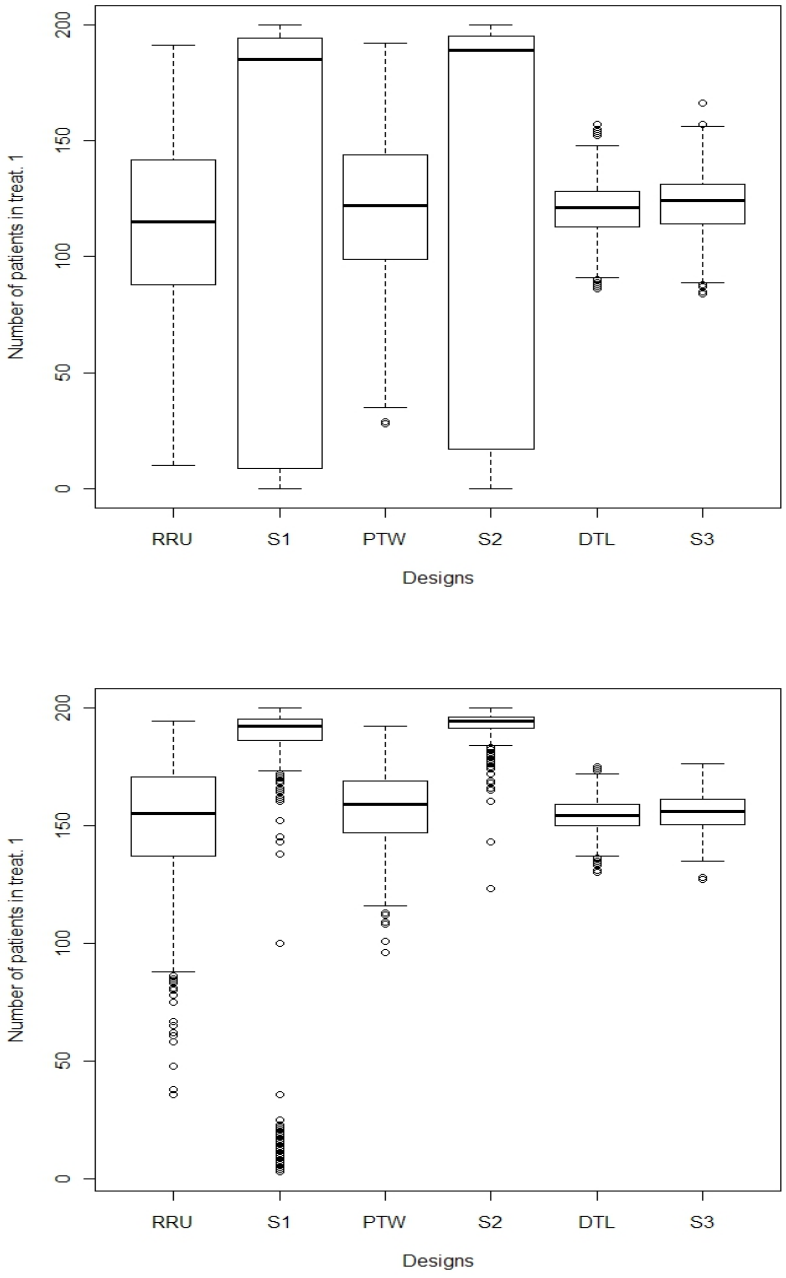

Figure 2.3: Comparison of the three scenarios presented (S1, S2, S3) with their inspiring designs, RRU, PTW and DTL. 1000 replications of the trial for 200 patients and success probabilities $p_1 = 0.9$, $p_2 = 0.8$ and $p_1 = 0.9$, $p_2 = 0.5$

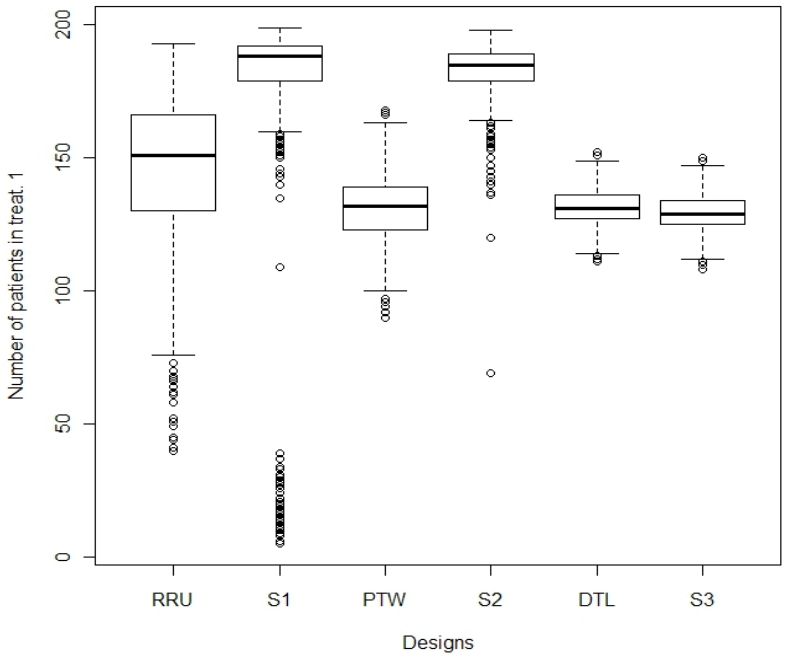

Figure 2.4: Comparison of the three scenarios presented (S1, S2, S3) with their inspiring designs, RRU, PTW and DTL. 1000 replications of the trial for 200 patients and success probabilities $p_1 = 0.7$ and $p_2 = 0.4$

Observing the figures, we can get a good idea about the performance of the three scenarios. In S1 and S2, the rule allocates most of the patients to the best treatment than their inspiring rules, so we can say that they are more ethical. Nevertheless, we can see a lot of variance and many extreme cases. For S3, we can observe that the behavior is very similar to the DTL rule. We already know that the limit of the proportion of allocations in treatment 1 is the same, but the similarity of the box suggests a similarity in the variance. As the DTL rule has been proven competitive, we have evidence that S3 could have good properties, so we have to confirm it theoretically.

The next table is built in the spirit of the tables of [62], comparing an ethical measure and an inferential one. n represents the number of patients that generate a simulated power of 90% for the test $\frac{\hat{p}_{n,1}-\hat{p}_{n,2}-(p_1-p_2)}{\sqrt{\frac{p_1 q_1}{N_{n,1}}+\frac{p_2 q_2}{N_{n,2}}}}$ when the complete randomization design is

applied. Then, the simulated power and the mean number of failures (with the standard deviation between round brackets) are obtained for the $S1$, $S2$ and $S3$ designs and their classical related rules. Observe that the expected number of failures for the S3 design is exactly obtained from.

Table 2.1: Simulated power and expected number of failures (standard deviation). 5.000 replications.

			RRU		S1	
p_1	p_2	n	Power	Failures	Power	Failures
0.9	0.8	532	83	75 (12.4)	17	75 (26.1)
0.9	0.5	48	80	12 (3.9)	52	9 (4.9)
0.7	0.4	108	82	43 (6.6)	41	39 (9.4)
			PTW		S2	
p_1	p_2	n	Power	Failures	Power	Failures
0.9	0.8	532	87	75 (8.9)	20	68 (22.9)
0.9	0.5	48	84	11 (3.1)	39	7.4 (2.9)
0.7	0.4	108	88	45 (5.6)	43	36 (5.7)
			DTL		S3	
p_1	p_2	n	Power	Failures	Power	Failures
0.9	0.8	532	89	73 (7.9)	89	72.04
0.9	0.5	48	86	11 (2.6)	87	11.41
0.7	0.4	108	87	44 (5.4)	87	44.30

Note that $S1$ and $S2$ designs are competitive with, or slightly better than their related rules, RRU and PTW, from the point of view of the number of failures, but clearly inferior from the point of view of the power of the test statistic. This loss of power could be foreseen from the previous comparative graphic. In that graphic, we could observe that S1 and S2 have greater variance in the number of allocations to treatment 1. From the results in [32], the higher the variability of $N_{n,1}$ the smaller is the power of the test statistic, as is clearly seen in the table. On the other hand, S3 is again confirmed as a rule very similar to the

DTL rule, with almost similar values in both measures, power and expected number of failures.

In summary, we have presented three new response-adaptive designs, inspired in three already well known designs. Although they have a good ethical behavior, we can confirm that S1 and S2 are not adequate as clinical trials, due to their very bad inferential properties, low power and very high variability in the number of allocations. As the DTL rule is perceived as a competitive response-adaptive design, the S3 scenario seems promising and a deeper theoretical study appears necessary.

2.2 Klein urn design

S3 design was already presented in [41], like a stochastic process for physics applications and not like a clinical trial design. The urn model will be called Klein urn model in the sequel and the design, Klein urn design or Klein design. The procedure of the design is as follows. Patients arrive sequentially and their responses to treatments are known before the arrival of the next patient. For each patient, a ball is drawn. If the ball is of type 1, the patient receives treatment 1; otherwise, it receives treatment 2. If the treatment is successful, the ball is replaced in the urn; otherwise, the ball extracted is removed and a ball of the other type is added to the urn. Therefore, the number of balls of type 1 in the urn after the n-th replacement, $W_{n,1}$, increases one step if a treatment of type 2 is applied and it is a failure and decreases one step if treatment type 1 is applied and it is a failure; otherwise, the number of type 1 balls in the urn remains unchanged. Then, as was stated before, the evolution of the number of balls of type 1 in the urn, $\{W_{n,1}\}$, is a Markov chain with state space $E = \{0, ..., 2w\}$ and the following transition matrix $P = (p_{i,j})$,

$$p_{i,j} = \begin{cases} q_2(1 - \frac{i}{2w}), & j = i + 1; \\ p_1 \frac{i}{2w} + p_2(1 - \frac{i}{2w}), & j = i; \\ q_1 \frac{i}{2w}, & j = i - 1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 0, \ldots, 2w. \tag{2.16}$$

In the first section we have made a study of the processes associated to the urn design, like $\{W_{n,1}\}$ and $\{N_{n,1}\}$, for the three scenarios. In this section, as we are focused on the Klein urn, we are able to extend these results.

Following the notation presented in Chapter 1, let $\mathcal{F}_n$ be the natural σ-algebra of the design, with the past allocations and responses. Formally, $\mathcal{F}_n = \sigma(\mathcal{G}_n, \mathcal{Z}_n)$, being $\mathcal{G}_n$ the set of allocations until the n-th patient and $\mathcal{Z}_n$ the set of observed responses until the n-th patient. We also define R_n as the random variable determining the number of balls of type 1 added to the urn in the n-th step, which can be decomposed like

$$R_n = Z_{n,1}(1 - \delta_n) - Z_{n,2}\delta_n. \tag{2.17}$$

Note that

$$W_n = W_{n-1} + R_n. \tag{2.18}$$

Taking expectations in (2.17),

$$E[R_n] = q_2\,(1 - E[\delta_n]) - q_1 E[\delta_n]$$

From the allocation rule of the Klein urn design, the probability of assigning one treatment is the proportion of balls of this treatment in the urn. Using this we can derive $E[\delta_n] = E[E[\delta_n|\mathcal{F}_{n-1}]] = E[E[\delta_n|W_{n-1,1}]] = E[W_{n,1}/2w]$, so we have

$$E[R_n] = q_2(1 - E[W_{n-1,1}]/(2w)) - q_1 E[W_{n-1,1}]/(2w). \tag{2.19}$$

Squaring (2.17) and taking expectations, we get a similar expression for $E[R_n^2]$,

$$E[R_n^2] = q_2(1 - E[W_{n-1,1}]/(2w)) + q_1 E[W_{n-1,1}]/(2w). \tag{2.20}$$

To alleviate the notation, in what follows we denote $r = 1-((q_1+q_2)/(2w))$, $s = 2r-1$ and $\mu_N = q_2/(q_1+q_2)$. Note that $(w-1)/w < r < 1$, $(w-2)/w < s < 1$ and $\mu_W = 2w\mu_N$. We assume, without loss of generality, that $q_1 \leq q_2$. Then $\mu_W \geq w$, so the urn has in the limit more balls of the best treatment, if $q_1 \neq q_2$.

Proposition 2.2.1. *Assume that the Klein urn design is applied and $W_{0,1}$ is a generic random variable. Then,*

a)

$$E[W_{n,1}] = \mu_W + r^n d_0$$

where $d_0 = (E[W_{0,1}] - \mu_W)$,

b)

$$\begin{aligned} Var[W_{n,1}] &= \frac{2wq_1q_2}{(q_1+q_2)^2}(1-s^n) + d_0\frac{(q_1-q_2)}{q_1+q_2}(s^n-r^n) + d_0^2(s^n-r^{2n}) \\ &+ s^n Var[W_{0,1}]\ . \end{aligned}$$

Proof. The a) part is proven in Remark 2.1.4, now considering the most generic case of $W_{0,1}$ as a general random variable. For the variance we observe that, from (2.18),

$$Var[W_{n,1}] = Var[W_{n-1,1}] + Var[R_n] + 2Cov(W_{n-1,1},\, R_n).$$

From (2.17), we have that

$$Cov(W_{n-1,1},\, R_n) = E[W_{n-1,1}Z_{n,2}] - E[W_{n-1,1}(Z_{n,1}+Z_{n,2})\delta_n] - E[W_{n-1,1}]E[R_n]$$

Using the independence of the responses and (2.20),

$$\begin{aligned} &Cov(W_{n-1,1},\, R_n) = \\ &= q_2E[W_{n-1,1}] - \frac{q_1+q_2}{2w}E[W_{n-1,1}^2] - q_2E[W_{n-1,1}] + \frac{q_1+q_2}{2w}E^2[W_{n-1,1}] \\ &= -\frac{q_1+q_2}{2w}(E^2[W_{n-1,1}] - E[W_{n-1,1}^2]) = (r-1)Var[W_{n-1,1}]. \end{aligned} \tag{2.21}$$

Therefore

$$Var[W_{n,1}] = sVar[W_{n-1,1}] + Var[R_n],$$

and solving the recurrence,

$$Var[W_{n,1}] = \sum_{k=1}^{n} s^{n-k}Var[R_k] + s^nVar[W_{0,1}]. \tag{2.22}$$

So that $\{Var[W_{n,1}]/s^n\}$ increases with n. Moreover, combining (2.19) and (2.20) we have that,

$$Var[R_k] = q_2 - q_2^2 + \left[\frac{q_1 - q_2 + 2q_2(q_1+q_2)}{2w}\right] E[W_{k-1,1}] - \left[\frac{(q_1+q_2)}{2w}\right]^2 E^2[W_{k-1,1}]$$

From (a), we know that $E[W_{k-1,1}] = \mu_W + r^{k-1}d_0$, so we get

$$Var[R_k] = \frac{2q_1q_2}{q_1+q_2} + \frac{d_0}{2w}r^{k-1}\left[(q_1-q_2) - \frac{d_0}{2w}(q_1+q_2)^2 r^{k-1}\right] \tag{2.23}$$

Putting (2.23) in (2.22), b) follows. □

Remark 2.2.1. From the expression of Proposition 2.2.1 a), we can easily see the monotone convergence of $\{E[W_{n,1}]\}$ to μ_W. As $\{W_{n,1}\}$ is a Markov chain, this convergence is exponential which is also easily identifiable due to the appearance of a geometric progression with $|r| < 1$. Besides this, if $E[W_{0,1}] > \mu_W$, the sequence is strictly decreasing and if $E[W_{0,1}] < \mu_W$ the sequence is strictly increasing.

We have already seen the strong relationship between the treatment 1 allocation indicator, δ_n, and the number of balls of type 1 in the urn in the previous step, $W_{n-1,1}$. Using this relationship we get the next corollary, which gives exact results for each n for the $\{\delta_n\}$ process.

Corollary 2.2.1. *Assume that the Klein urn design is applied and $W_{0,1}$ is a random variable taking values in E. Then,*

a)

$$E[\delta_n] = \mu_N + r^{n-1}\left(\frac{d_0}{2w}\right)$$

b)

$$Var[\delta_n] = \left(\mu_N + r^{n-1}\frac{d_0}{2w}\right)\left(1 - \mu_N - r^{n-1}\frac{d_0}{2w}\right)$$

Proof. Observe that

$$E[\delta_n] = E[E[\delta_n|\mathcal{F}_{n-1}]] = E[E[\delta_n|W_{n-1,1}]] = \frac{E[W_{n-1,1}]}{2w} \tag{2.24}$$

for $n \geq 1$, and therefore, a) and b) follow straightforwardly, using the a) result from Proposition 2.2.1. □

The following result is a technical lemma, which is going to be useful during the proofs of this chapter and which connects the variance of variables with the covariances of their conditional expectation.

Lemma 2.2.1. *If X and Y are random variables and X is $\mathcal{F}$-measurable, then*

$$Cov(X,Y) = Cov(X, E[Y|\mathcal{F}])$$

Proof.

$$\begin{aligned} Cov(X, E[Y|\mathcal{F}]) &= E[XE[Y|\mathcal{F}]] - E[X]E[E[Y|\mathcal{F}]] \\ &= E[E[XY|\mathcal{F}]] - E[X]E[Y] = Cov(X,Y) \end{aligned}$$

□

Let $\boldsymbol{\delta}_n = (\delta_1, \delta_2, \ldots, \delta_n)$ be the vector of indicator variables of type 1 treatment. The components of the covariance matrix of $\boldsymbol{\delta}_n$, $\Sigma_{\boldsymbol{\delta}_n}$, can be expressed in terms of known processes as follows.

Proposition 2.2.2. *Assume that the Klein urn design is applied. The components of the covariance matrix of $\boldsymbol{\delta}_n$ satisfy the following relations*

$$Cov(\delta_i, \delta_j) = \begin{cases} Var[\delta_i], & i = j; \\ \dfrac{1}{4w^2} Var[W_{i-1,1}] + (r-1)Var[\delta_i], & j = i+1; \\ r^{j-i-1} Cov(\delta_i, \delta_{i+1}), & j > i+1. \end{cases}$$

Proof. Consider first that $j > i+1$, with $i = 1, \ldots, n$. From Lemma 2.2.1 and (2.24) we have

$$\begin{aligned} Cov(\delta_i, \delta_j) &= Cov(\delta_i, E[\delta_j | W_{j-1,1}]) \\ &= \frac{1}{2w}(E[\delta_i W_{j-1,1}] - E[\delta_i]E[W_{j-1,1}]) \, . \end{aligned}$$

Using the recursive relation (2.18),

$$\begin{aligned} Cov(\delta_i, \delta_j) &= \frac{1}{2w}(E[\delta_i W_{j-2,1}] + E[\delta_i R_{j-1}] - E[\delta_i]E[W_{j-2,1}] - E[\delta_i]E[R_{j-1}]) \\ &= Cov(\delta_i, \delta_{j-1}) + \frac{1}{2w} Cov(\delta_i, R_{j-1}) \, . \end{aligned} \tag{2.25}$$

As the response of the patients is independent of the previous history of allocations and responses, from the definition of R_{j-1}, $R_{j-1} = Z_{j-1,1}(1-\delta_{j-1}) - Z_{j-1,2}\delta_{j-1}$, we have

$$Cov(\delta_i,\, R_{j-1}) = -(q_2+q_1)Cov(\delta_i,\, \delta_{j-1})$$

and plugging this result in (2.25), we finally have

$$Cov(\delta_i,\, \delta_j) = rCov(\delta_i,\, \delta_{j-1}). \tag{2.26}$$

Developing the recurrence, we get the result,

$$Cov(\delta_i,\, \delta_j) = r^{j-i-1}Cov(\delta_i,\, \delta_{i+1}).$$

For the case of $j = i+1$, proceeding as in (2.25) - (2.26) we obtain

$$\begin{aligned} Cov(\delta_i,\, \delta_{i+1}) &= \frac{1}{2w}(E[\delta_i W_{i-1,1}] - E[\delta_i]E[W_{i-1,1}]) - (1-r)Var[\delta_i] \\ &= \frac{1}{4w^2}Var[W_{i-1,1}] + (r-1)Var[\delta_i] \end{aligned}$$

and the result follows. □

The previous results were focused on the number of type 1 balls in the urn, $W_{n,1}$, and the treatment 1 allocation indicators, δ_n. Another relevant process in a clinical trial is the number of allocations to treatment 1 until the n-th patient, $N_{n,1}$. Obviously, $N_{n,1} = \sum_{i=1}^{n}\delta_i$. Using this relation and (2.24), we can obtain exact and asymptotic results for this process.

Proposition 2.2.3. *Assume that the Klein urn design is applied. For each $n > 0$, we have*

a)

$$E\left[N_{n,1}\right] = n\mu_N + \frac{(1-r^n)}{(1-r)}\left(\frac{E[W_{0,1}]}{2w} - \mu_N\right). \tag{2.27}$$

b)

$$Var[N_{n,1}] = \sum_{i=1}^{n} Var[\delta_i] + \frac{2}{1-r}\sum_{i=1}^{n-1}\left((1-r^{n-i})Cov(\delta_i, \delta_{i+1})\right). \tag{2.28}$$

Proof. In Remark (2.1.4) a) has been proven. To prove b), observe that

$$Var[N_{n,1}] = Var\left[\sum_{i=1}^{n} \delta_i\right] = \mathbf{1}^t \Sigma_{\boldsymbol{\delta}_n} \mathbf{1},$$

where $\mathbf{1}$ is the column vector of ones and $\mathbf{1}^t$ is its transpose. From Proposition 2.2.2, b) follows straightforwardly.

□

A strong law and a central limit theorem can be obtained for the allocation process $\{N_{n,1}\}$.

Proposition 2.2.4. *Let $\{N_{n,1}\}$ be the allocation process of treatment 1 when the Klein urn design is applied. Then,*

a)

$$\frac{N_{n,1}}{n} \to \mu_N, \quad a.s.$$

b)

$$\sqrt{n}\left(\frac{N_{n,1}}{n} - \mu_N\right) \to N(0,\ \sigma_N^2)$$

where $\sigma_N^2 = \dfrac{q_1 q_2 (p_1 + p_2)}{(q_1 + q_2)^3}$.

Proof. Note that $E[\delta_n | \mathcal{F}_{n-1}] = W_{n-1,1}/2w$. We can build a martingale $\{M_n\}$ such that

$$M_n = \sum_{k=1}^{n} (\delta_k - E[\delta_n | \mathcal{F}_{n-1}]) = \sum_{k=1}^{n} \left(\delta_k - \frac{W_{k-1,1}}{2w}\right)$$

and $\{(1/2w)\sum_{k=0}^{n-1} W_{k,1}\}$ is the compensator of the allocation process $\{N_{n,1}\}$. As the Markov chain $\{W_{n,1}\}$ is ergodic, we are in the conditions of Theorem 1 in [6]. Therefore, we have that a) follows and $\sqrt{n}(\frac{N_{n,1}}{n} - \mu_N)$ converges in distribution to a random variable with normal distribution and zero mean. We only need to obtain the variance of this limit distribution. That is, we have to calculate $lim_{n\to\infty} Var[N_{n,1}]/n$.

Proposition 2.2.3, b) gives an explicit expression of $Var[N_{n,1}]$,

$$Var[N_{n,1}] = \sum_{i=1}^{n} Var[\delta_i] + \frac{2}{1-r} \sum_{i=1}^{n-1} \left((1 - r^{n-i}) Cov(\delta_i, \delta_{i+1})\right).$$

The expression is in terms of $Var[\delta_n]$ and $Cov(\delta_n, \delta_{n+1})$, so we only need the limits of these quantities. From Proposition 2.2.1 and the fact that the stationary distribution of $\{W_{n,1}\}$ is binomial, we have

$$Var[\delta_n] \to \frac{q_1 q_2}{(q_1+q_2)^2}.$$

From Proposition 2.2.2 we have

$$Cov(\delta_n, \delta_{n+1}) = \frac{1}{4w^2} Var[W_{n-1,1}] + (r-1)Var[\delta_n],$$

and taking limits,

$$Cov(\delta_n, \delta_{n+1}) \to \frac{1}{2\omega} \frac{q_1 q_2}{(q_1+q_2)^2} + (r-1)\frac{q_1 q_2}{(q_1+q_2)^2}.$$

Then, applying Toeplitz lemma in the expression of $Var[N_{n,1}]/n$, b) follows. □

Remark 2.2.2. It is worth highlighting that these asymptotic limits are the same as those obtained when the drop-the-loser rule is applied. The proportion of allocations to treatment 1 up to the n-th patient, $N_{n,1}/n$, converges to $\frac{q_2}{q_1+q_2}$, the urn allocation, in both designs. In the central limit theorem, the asymptotic variance is also the same, $\sigma_N^2 = \frac{q_1 q_2 (p_1+p_2)}{(q_1+q_2)^3}$, which is the minimum variance that could be reached with the urn limit allocation, as was shown in [34].

Remark 2.2.3. In order to obtain the asymptotic variance σ_W^2 in (2.6), reasoning as in Proposition 2.2.4, we solve $\lim_{n\to\infty} Var[\sum_{k=1}^n W_{k,1}]/n$. We decompose the variance

$$Var\left[\sum_{k=1}^n W_{k,1}\right] = \sum_{k=1}^n Var[W_{k,1}] + 2\sum_{k>j} Cov(W_{k,1}, W_{j,1})$$

We know that $W_{k,1} = W_{k-1,1} + R_k$, so, for each $k > j$, combining (2.17) and Lemma 2.2.1, we get

$$Cov(W_{k,1}, W_{j,1}) = Cov(W_{k-1,1} + R_k, W_{j,1}) = r^{k-j} Var(W_{j,1}).$$

If we introduce this recurrence in the previous sum,

$$\begin{aligned} Var\left[\sum_{k=1}^{n} W_{k,1}\right] &= \sum_{k=1}^{n} Var[W_{k,1}] + 2\sum_{j=1}^{n} r^{k-j} Var[W_{j,1}] \\ &= \sum_{k=1}^{n} Var[W_{k,1}] + 2\sum_{j=1}^{n}\sum_{k=j+1}^{n} r^{k-j} Var[W_{j,1}] \\ &= \sum_{k=1}^{n} Var[W_{k,1}] + \frac{2r}{1-r}\sum_{j=1}^{n} Var[W_{j,1}](1-r^{n-j}). \end{aligned}$$

From Proposition 2.2.1 b) we have a closed expression for $Var[W_{k,1}]$. So, taking limits in $Var[\sum_{k=1}^{n} W_{k,1}]/n$, we have $\sigma_W^2 = 4w^2\sigma_N^2(1+r)/(p_1+p_2)$.

Corollary 2.2.2. *Assume that the Klein urn design is applied and $q_1, q_2 \in (0,1)$ with $q_1 \neq q_2$. Then, there exists a value $i_0 \geq 2$ such that, for any pair of indices $j > i > i_0$,*

a) If $q_1 + q_2 > 1$, then $Cov(\delta_i, \delta_j) < 0$.

b) If $q_1 + q_2 < 1$, then $Cov(\delta_i, \delta_j) > 0$.

Proof. Without loss of generality, let us assume that $q_2 > q_1$. As $q_1, q_2 \in (0,1)$, $Var[\delta_i] > 0$, for any $i \geq 2$. From Proposition 2.2.2 we have that $Cov(\delta_i, \delta_{i+1})$ and $Cov(\delta_i, \delta_j)$ with $j > i$ have the same sign, and they are positive when

$$\frac{Var[W_{i-1,1}]}{Var[\delta_i]} > 2w(q_1+q_2). \tag{2.29}$$

As the sequence $\left\{\frac{Var[W_{i-1,1}]}{Var[\delta_i]}\right\}$ converges to $2w$, when $i \to \infty$, from (2.29) we get a) and b). □

Remark 2.2.4. If $q_1 = q_2 = q$ and $W_{0,1} = w$, then, $d_0 = 0$ and $Var[W_{0,1}] = 0$ in Proposition 2.2.1 b). Therefore, $Var[W_{i,1}] = w(1-s^i)/2$ and $Var[\delta_i] = 1/4$, so that, the sequence $\{Var[W_{i-1,1}]/Var[\delta_i]\}$ becomes $\{2w(1-s^{i-1})\}$. For any $w \geq 2$, this sequence strictly increases to $2w$. Therefore, if $q > 1/2$, $Cov(\delta_i, \delta_j) < 0$ for any $i < j$. If $q < 1/2$, the sequence $\{Var[W_{i-1,1}]/Var[\delta_i]\}$ will eventually be greater than $2w2q$, and therefore, there exists i_0, namely $i_0 = log(1-2q)/log(s)$, such that, $Cov(\delta_i, \delta_j) > 0$ for any $i_0 < i < j$. If i_0 is a positive integer, then $Cov(\delta_{i_0}, \delta_j) = 0$, for any $j > i_0$.

2.3 Performance of the Klein design

In the previous section a probabilistic analysis of the different processes associated to the urn has been made when the Klein has been applied for the allocations. Now, we present different properties about the performance of the design, focusing more on a practical aspect. As we have presented a new design, we should compare it with other designs in the literature and study its performance properties.

In this section we obtain explicit formulae for the main operating characteristics of the Klein design, such as the expected treatment failures, power with the Wald test, selection bias and accidental bias. As we have seen, the Klein design and the DTL design have the same limit allocation and the same asymptotic variance in the proportion of allocations. So that, a similar behavior for their operating characteristics should be expected. However, the Klein design has an important advantage which is that, for each n, exact formulae have been obtained for the expectation and the variance of the number of patients in each treatment.

2.3.1 Expected number of failures and power

In the specialized literature, some comparative studies have been made among outstanding response-adaptive designs in order to establish a preference for a particular procedure according to their degree of compromise among ethics, randomness and inferential accuracy. As mentioned in Chapter 1, the DBCD2 design and the DTL design have been found competitive due to their good operating characteristics, see, for instance, [62] and [24].

In [62] the power of the usual Wald test and the expected number of failures are obtained, via simulation, for the DTL design (Table 2) and for the DBCD2 design (Table 3), for the number of patients, n, that provide a power of 90% for the complete randomized design. Using the previous notation, the expected number of failures can be expressed as follows,

$$E\left[\sum_{i=1}^{n}(\delta_i Z_{n,1} + (1-\delta_i)Z_{n,2})\right] = (q_1 - q_2)E[N_{n,1}] + nq_2,$$

and we can obtain, using Proposition 2.2.3, this expected value for any initial state $W_{0,1}$.

Proposition 2.2.3 implies that $N_{n,1} \to \infty$ and $N_{n,2} \to \infty$ when the Klein design is applied. Then, the conditions of Theorem 1 in [11] are fulfilled, and a central limit theorem holds for the proportion of successes of each treatment. Therefore, the Wald test can be used for large sample sizes. These results are reproduced in our Table 2.2 in order to compare them with the expected number of failures for the Klein design (which is calculated assuming that $W_{0,1} = w$, and the power, obtained for the Klein design in the same simulation conditions, that is, 10,000 replications ($\alpha = 0.05$, two-sided). Negligible differences appear between the behavior of the DTL and the Klein design, so that Klein design would be preferred to the DBCD2 design in the same situations in which the DTL rule would be preferred.

Note also that the power of the usual Wald test when the Klein design is used is similar to the power when the DTL design is used, and it is within 1% of complete randomization.

Table 2.2: Simulated expected number of failures (standard deviation) for the DBCD2 and DTL designs, exact values for the Klein design, and simulated power (10,000 replications).

			DBCD2		DTL		Klein	
p_1	p_2	n	Power	Failures	Power	Failures	Power	Failures
0.9	0.3	24	91	8 (1.7)	90	7 (1.8)	90	7.71
0.9	0.5	50	91	13 (2.6)	89	12 (2.6)	88	11.81
0.9	0.7	162	90	31 (4.8)	89	27 (4.6)	88	26.71
0.9	0.8	532	91	79 (8)	89	73 (8)	89	72.04
0.7	0.3	62	90	28 (3.5)	89	27 (4.1)	89	27.57
0.7	0.5	248	90	97 (7.5)	89	93 (8.0)	89	93.62
0.5	0.4	1036	90	567 (16)	89	565 (16)	90	565.17
0.3	0.1	158	90	122 (5.4)	90	124 (5.3)	89	124.58
0.2	0.1	532	90	448 (9)	90	451 (8)	90	450.67

2.3.2 Selection Bias

One of the main advantages of the use of random allocation rules in the assignment of patients is the impossibility to accurately guess the treatment that is going to be applied to the next patient. Even so, in unmasked clinical trials and due to the sequential arrival of patients, the researcher could have a guess of which treatment is more likely to be assigned and this could consciously or unconsciously affect the recruitment process and bias the study. Therefore, a natural way to reduce selection bias is to increase the randomness of the design, avoiding deterministic or skewed allocations. For instance, the best allocation rule in terms of randomness is complete randomization, because all the allocations have probability equal to 1/2. As we move away from this probability, randomness is lost and the design is more vulnerable to selection bias. In response-adaptive designs, the allocation probabilities are skewed because this is one of the goals of the design, but we can still think about selection bias.

In [12], a measure of randomness is presented, which is based on the in mean distance between the proportion of guesses of the next treatment to be applied, with an optimal guessing strategy, and the maximum uncertainty, which is equal probability. We present a measure based on that, called average selection bias. Let J_k be the indicator variable of the event "*correct guess of the treatment applied to the kth patient*". Observe that the best strategy is to predict the treatment with more balls in the urn. Therefore, $E[J_k] = E[max(W_{k-1,1}, 2w - W_{k-1,1})]/(2w) = (w + E[|W_{k-1,1} - w|])/(2w)$. We consider the average selection bias up to the epoch n given by

$$\begin{aligned} \beta_n &:= \frac{1}{n}\sum_{k=1}^{n} E[J_k] \\ &= \frac{1}{2} + \frac{1}{2wn}\sum_{k=1}^{n} E[|W_{k-1,1} - w|] \\ &= \frac{1}{2} + \frac{1}{2wn}\sum_{k=1}^{n}\left(\sum_{i=1}^{w} i(P(W_{n,1} = w + i) + P(W_{n,1} = w - i))\right) . \end{aligned} \tag{2.30}$$

Using that the stationary distribution of $\{W_{n,1}\}$ for the Klein urn model is binomial

with parameters $2w$ and $q_2/(q_1+q_2)$, we have from (2.30) that

$$\beta_n \to \beta_\infty := \frac{1}{2} + \frac{1}{2w}\left(\frac{l}{(l+1)^2}\right)^w \sum_{i=1}^{w} i \binom{2w}{w+i}(l^i + l^{-i}) \tag{2.31}$$

where $l = q_2/q_1$.

Remark 2.3.1. Note that β_∞ does not change if we define $l = q_1/q_2$. This expression is related with the absolute deviation of the binomial distribution (see [40] for a wider study about absolute deviations) and it can be expressed, using the Gaussian hypergeometric function, as

$$\beta_\infty = \frac{1}{l+1} + \frac{1}{w}\binom{2w}{w+1}\frac{l^{w+1}}{(l+1)^{2w}}\,{}_2F_1(2, 1-w, w+2, -l)\,.$$

In Table 2.3 the average selection bias of the Klein design is presented, β_n, for some combinations of values q_1 and q_2, and for $n = 50, 200$. These quantities are calculated assuming that $W_{0,1} = w$, and using formula (2.30), where the exact values of the n-step probabilities are those given in [41]. This table also provides the average and standard deviation of β_n for the DTL and DBCD2 designs, obtained by simulation (10,000 replications). Again, the DTL and the Klein design perform in a quite similar way. We observe that the biased coin design (DBCD2 design) is less predictable for small ratios l .

Table 2.3: Average selection bias (s.d.) for DTL and DBCD2 designs (10.000 replications), exact values for the Klein design.

			DTL		DBCD2		Klein		
l	q_1	q_2	50	200	50	200	50	200	∞
1	0.2	0.2	0.58(0.07)	0.59(0.04)	0.68(0.08)	0.63(0.04)	0.56	0.58	0.588
1	0.6	0.6	0.64(0.06)	0.64(0.03)	0.55(0.07)	0.54(0.04)	0.58	0.59	0.588
0.7	0.35	0.5	0.63(0.06)	0.64(0.03)	0.6(0.08)	0.59(0.05)	0.59	0.61	0.614
0.5	0.25	0.5	0.65(0.06)	0.67(0.03)	0.64(0.09)	0.66(0.06)	0.61	0.65	0.672
0.3	0.15	0.5	0.67(0.07)	0.73(0.04)	0.61(0.07)	0.62(0.05)	0.64	0.73	0.769

2.3.3 Accidental Bias

When a clinical trial is planned, the covariates that could influence in the outcome of the patient are included in the design and in the allocation rule. However, it is impossible to

predict all the possible effects in the patients with accuracy, so, sometimes, there could be some important covariates that have not been taken into account. Accidental bias is defined as the effect of a covariable that has not been taken into account and which has influence on the response. An imbalance in those covariates could produce a bias in the treatment effect estimation. The randomization process by itself mitigates this problem, but in asymptotic terms. In small samples there could be large imbalances.

On the other hand, the approach to accidental bias is different in treatment-adaptive designs and response-adaptive designs. Treatment-adaptive designs pursue balance between the two treatments, and also, a balance in the covariables is desirable. In response-adaptive designs, the limit of the allocations is skewed to the best treatment, so the same distribution of covariate levels among treatmens is pursued.

One natural wish is that the covariate effect is similar in the two treatments, so a way to measure the accidental bias could be the unbalance of the covariate effect between the two treatments. If we denote the covariate as a random variable H, a measure of covariate unbalance (see [63]) could be

$$\Delta F_n = \frac{\sum_{i=1}^{n} \delta_i H_i}{N_{n,1}} - \frac{\sum_{i=1}^{n} (1-\delta_i) H_i}{N_{n,2}} \,,$$

where H_i is the value of covariate H in the i-th patient. If we assume that $\{H_i\}$ is a sequence of independent and identically distributed random variables with variance σ^2, then

$$E[\Delta F_n] = 0 \quad \text{and} \quad Var[\Delta F_n] = \frac{\sigma^2}{n} E\left[\frac{n^2}{N_{n,1}(n-N_{n,1})}\right] .$$

If the expectation $E\left[\frac{n^2}{N_{n,1}(n-N_{n1,})}\right]$ is uniformly bounded by a constant, we can use the Markov property, and we have that, for any value $\varepsilon > 0$,

$$P(|\Delta F_n| > \varepsilon) \to 0, \qquad \text{when } n \to \infty \,. \tag{2.32}$$

A different approach is possible. As we have seen, the conditions of Theorem 1 in [11] are fulfilled for the Klein design. Then, we have a strong law for $\sum_{i=1}^{n} \delta_i F_i / N_{n,1}$ and

$\sum_{i=1}^{n}(1-\delta_i)F_i/N_{n,2}$. Therefore, $\Delta F_n \to 0$ a.s. and, in particular, (2.32) holds.

In order to evaluate the speed of convergence, we simulate the probability of unbalance for the Klein design, the DTL design and the DBCD2 design. Results are presented in Table 2.4. We take $n = 50, 200, 1000$ and $(p_1, p_2) = (0.8, 0.8), (0.8, 0.5), (0.8, 0.2)$. The sequence $\{Fi\}$ are independent Bernoulli random variables with success probability 0.3. The average of the proportion of trials with the standard deviation between brackets, in 1000 sets of 1000 replications of the trial, for which $|\Delta F_n| > 0.1$ is computed for each combination of size and success probabilities. We see that, in spite of the convergence to 0, the probability of unbalance is not negligible for small values of n. Observe that the Klein design has slightly smaller values for the measure of covariate imbalance.

Table 2.4: Simulation of $P(|\Delta F_n| > \varepsilon)$, $\varepsilon = 0.1$, average of the frequency (standard deviation below) for DTL, DBCD2 and Klein designs (1000 sets of 1000 replications of the design).

n	50			200			1000		
p_1, p_2	DTL	DBCD2	Klein	DTL	DBCD2	Klein	DTL	DBCD2	Klein
0.8, 0.8	0.450	0.463	0.445	0.125	0.127	0.125	0.0005	0.0006	0.0006
	(0.039)	(0.040)	(0.040)	(0.016)	(0.017)	(0.016)	(0.0007)	(0.0008)	(0.0008)
0.8, 0.5	0.498	0.493	0.458	0.166	0.166	0.1539	0.0018	0.0019	0.0017
	(0.036)	(0.038)	(0.039)	(0.018)	(0.0183)	(0.018)	(0.0014)	(0.0014)	(0.0013)
0.8, 0.2	0.548	0.534	0.484	0.220	0.216	0.197	0.0058	0.0059	0.0052
	(0.040)	(0.042)	(0.036)	(0.020)	(0.020)	(0.020)	(0.0025)	(0.0026)	(0.0024)

The previous approach fails if the independence of the covariate sequence cannot be guaranteed. A different approach is followed in [19], where a linear model for the response of each patient, y_k, is assumed:

$$y_k = \mu + \alpha\tilde{\delta}_k + \beta F_k + \varepsilon_k,\ k = 1, \ldots, n \tag{2.33}$$

where $\tilde{\delta}_k = 2\delta_k - 1$, F_k is the measure of the covariate and ε_k is the error.

As in [70], we denote $u_k = \beta F_k + \varepsilon_k$, the total error in the mean treatment difference

when the covariate is not included, then

$$(2N_{n,1}(n-N_{n,1}))^{-1}[n\sum_{k=1}^{n}u_k\tilde{\delta}_k]\,. \tag{2.34}$$

where it is assumed, without loss of generality, that $\sum_{k=1}^{n}u_k=0$. So that, if $\mathbf{u}_n=(u_1,\ldots,u_n)$ and $\Sigma_{\tilde{\boldsymbol{\delta}}_n}$ is the covariance matrix of $\tilde{\boldsymbol{\delta}}_n=(\tilde{\delta}_1,\ldots,\tilde{\delta}_n)$, the error variance of $\hat{p}_1-\hat{p}_2$ is proportional to $\sigma^2_{\mathbf{u}_n}:=\mathbf{u}_n'\Sigma_{\tilde{\boldsymbol{\delta}}_n}\mathbf{u}_n$. The maximum of $\sigma^2_{\mathbf{u}_n}$ is a plausible measure of the accidental bias (see [19], [70] and [63]). Besides, for each $\mathbf{x}\in\mathbf{R}^n$

$$\frac{\mathbf{x}^t\Sigma_{\tilde{\boldsymbol{\delta}}_n}\mathbf{x}}{\mathbf{x}^t\mathbf{x}}\leq\lambda_{max}(\Sigma_{\tilde{\boldsymbol{\delta}}_n}) \tag{2.35}$$

where $\lambda_{max}(A)$ represents the maximum eigenvalue of the square matrix A. So that, under the normalization condition $\sum_{k=1}^{n}u_k^2=1$, we have that $\sigma^2_{\mathbf{u}_n}\leq\lambda_{max}(\Sigma_{\tilde{\boldsymbol{\delta}}_n})$, and then the maximum eigenvalue of $\Sigma_{\tilde{\boldsymbol{\delta}}_n}$ becomes a bound for the accidental bias.

Observe that $\Sigma_{\tilde{\boldsymbol{\delta}}_n}=4\Sigma_{\boldsymbol{\delta}_n}$, so that, in what follows, we focus on the study of the eigenvalues of $\Sigma_{\boldsymbol{\delta}_n}$. From Proposition 2.2.2 and Corollary 2.2.2 we have a good theoretical knowledge of the components of this covariance matrix.

The Klein urn model could always be in a stationary regime. It would be sufficient that the distribution of the initial state $W_{0,1}$ were binomial with parameters $2w$ and $q_2/(q_1+q_2)$. But this is unrealistic, because the probabilities q_1 and q_2 are unknown. Nevertheless, as it was done in [19] for the Efron design, we are going to obtain the maximum eigenvalue for the Klein urn model when it is in stationary regime. Let

$$\rho_h:=\lim_{k\to\infty}Cov(\delta_h,\delta_{h+k}),$$

then, from Proposition 2.2.2 we have that

$$\rho_k=\begin{cases}\dfrac{q_1q_2}{(q_1+q_2)^2}, & k=0\,;\\ r^{k-1}\rho, & k\geq 1\,,\end{cases}$$

where $\rho = \dfrac{q_1 q_2 (1-(q_1+q_2))}{2w(q_1+q_2)^2}$. As $\sum |\rho_k| < \infty$, the spectral density, $f(s)$, $s \in [0,\, \pi]$, is

$$\begin{aligned} f(s) &= \sum_{k=-\infty}^{\infty} \rho_k e^{-isk} \\ &= \rho_0 + 2\rho \sum_{k=1}^{\infty} r^{k-1} \cos(sk) \\ &= \rho_0 - \frac{2\rho}{r} + \frac{2\rho}{r} \frac{1 - r\cos(s)}{1 - 2rcos(s) + r^2}\,, \end{aligned} \tag{2.36}$$

where the last equality follows from exercise 189 of chapter XII in [42]. Besides, we have

$$f(0) = \sigma_N^2 \qquad f(\pi) = \rho_0 \left(\frac{4w - (p_1 + p_2)}{4w - (q_1 + q_2)} \right).$$

From the analysis of the derivative of $f(s)$ we have that, in the stationary regime,

a) If $q_1 + q_2 < 1$, $f(s)$ is strictly decreasing. Then σ_N^2 is an upper bound for the maximum eigenvalue.

b) If $q_1 + q_2 = 1$, $f(s) = \rho_0$ for any value s. Then ρ_0 is the maximum eigenvalue.

c) If $q_1 + q_2 > 1$, $f(s)$ is strictly increasing. Then $f(\pi)$ is an upper bound for the maximum eigenvalue.

These conditions can be seen as a hint of what happens in the general case.

Remark 2.3.2. A bound that is valid for the stationary case might not be valid in general. This is the case in the Efron design, for which the maximum eigenvalue in the stationary analysis is smaller than the value of the left hand side in (2.35) when $n = 2$ and $\mathbf{x}^t = (1/\sqrt{2}, -1/\sqrt{2})$, as stated in [70]. This problem has been revisited in [46] where they conjecture that the maximum eigenvalue of the covariance matrix does not depend on n.

Intensive numerical calculations, suggest that the following conjecture, which has been proven in the stationary case, is also valid in the general case:

Conjecture: When $q_1+q_2 < 1$, the maximum eigenvalue grows with n towards σ_N^2 given in Proposition 4. Otherwise, it grows with n towards a constant value $a := a(p_1,\, p_2,\, w)$ for which a closed form expression has not been found.

Table 2.5: Maximum eigenvalue of covariance matrix, Σ_{δ_n}, for different values of n and probability pairs

q_1	q_2	100	500	1000	2000	∞
0.1	0.3	0.340713	0.693317	0.735494	0.746441	f(0)=0.75
0.1	0.5	0.257493	0.313943	0.321786	0.323538	f(0)=0.324074
0.3	0.5	0.304317	0.348628	0.350836	0.351383	f(0)=0.351562
0.3	0.7	0.261711	0.261711	0.261711	0.261711	
0.5	0.7	0.263667	0.263667	0.263667	0.263667	

Chapter 3

Exact Randomization Based Inference

As stated in Chapter 1, the procedure to recruit patients in a clinical trial may not be a random sampling and, so, there is no formal basis to accept the population model. In this context, the randomization procedure gives us a basis for inference by means of a randomization test. The null hypothesis of a randomization test is that the two treatments are equivalent, so the assignment process does not affect the patients' responses and these responses are considered fixed. The way to compute a p-value with a randomization test is to sum the probabilities of the treatment allocation permutations which lead to a value of the test statistic more extreme than the experimental one. This exact calculation requires the enumeration of all the permutations and computing the probability for each one, consistent with the allocation rule and the set of responses, which can be computationally unfeasible or very time consuming. In order to avoid all the enumeration, shorten the calculations and make the problem solvable, three ways have been presented in the specialized literature: the asymptotic distribution of the test statistic, Monte Carlo estimations of the p-values, and computational algorithms. In this chapter a recurrence algorithm to compute the exact distribution of two test statistics is presented, valid for a

wide family of response-adaptive designs, which is named $\Re$ family. Some designs in the literature that are included in the family are presented. With such designs, a simulation study is done, comparing different methods to obtain p-values and using these methods to make a power study. Finally, the algorithm is adapted to the DTL design, which is not in the $\Re$ family.

3.1 Exact Randomization Based Inference (RBI) for response-adaptive designs

Let Y_n be the response of the n-th patient and let $A_n = \{Y_1 = a_1, Y_2 = a_2, \ldots, Y_n = a_n\}$ be the set of responses of the first n patients. These responses are binary and assuming the randomization model, deterministic. Thus, all the processes in this chapter are conditioned to the particular set A_n. As the responses are dichotomous, $a_i = 1$ in case of success and $a_i = 0$ in case of failure. So, being the responses deterministic, the only random process is the allocation process, $\{\tilde{\delta}_i\}$, which is the sequence of centered allocation indicators. Let $\tilde{\delta}_n = 1$, if the patient is allocated in treatment 1, and $\tilde{\delta}_n = -1$ if it is allocated in treatment 2. It can be linked with the usual indicators through the linear transformation $\tilde{\delta}_n = 2\delta_n - 1$. In what follows, it is assumed that patients are randomized between treatments following a response-adaptive design, then the allocation process, $\{\tilde{\delta}_n\}$, depends on the past allocations and on the past deterministic responses.

The choice of a suitable test statistic is fundamental in any inferential procedure. Most of the nonparametric tests use linear rank tests and in the case of permutation tests, this is the usual procedure. In our case, we are going to consider a version of a linear rank test statistic for binary responses. Consider now the test-statistic

$$S_n = \sum_{i=1}^{n} a_i \tilde{\delta}_i \ , \tag{3.1}$$

which represents the difference between successes and failures in both treatments. S_n is the randomization test statistic for studying the equality of treatments. This choice

responds to the arguments given in section 2.5 in [56], where test statistics based on sum of responses are advised.

As stated in section 1.3, unconditional and conditional randomization tests are considered. The former implies that, given a number of patients n and a significance level α, the null hypothesis is rejected when the experiment gives a value s for the statistic S_n such that $P(|S_n| > s) < \alpha$. The latter behaves in a similar way but if the experiment has an imbalance in the number of patients $\Delta_n = N_{n,1} - N_{n,2} = t$ patients, the null hypothesis is rejected when $P(|S_n| > s | \, \Delta_n = t) < \alpha$.

The exact distribution of the test S_n is unknown and it must be computed enumerating all the possible permutations and calculating the probability of each arrangement. But the number of permutations increases very fast and even for moderate sizes of n it is computationally unfeasible. In [31] and [50] several algorithms to avoid all the calculations were presented, valid for a family of non response-adaptive designs. Here, we present a recursive algorithm to compute the exact distribution of the statistic S_n in a quite general framework. Our target now is to obtain an algorithm for finding the exact value of $P(S_n = s)$ for any value s in the interval $[-n, n]$.

Observe that

$$P(S_n = s) = \sum_{i=-n}^{n} P(S_n = s, \, \Delta_n = i) \tag{3.2}$$
$$P(\Delta_n = t) = \sum_{k=-n}^{n} P(S_n = k, \, \Delta_n = t)$$

and therefore, if $P(\Delta_n = t) > 0$,

$$P(S_n = s | \Delta_n = t) = \frac{P(S_n = s, \, \Delta_n = t)}{P(\Delta_n = t)}\,, \tag{3.3}$$

so that, in order to obtain the exact distribution for S_n via recurrence equations, in both cases, unconditional and conditional, is enough if a recurrence for $P(S_n = s, \, \Delta_n = t)$ is obtained. Besides this, this is also enough to obtain exact values for other test statistics that are transformations of the S_n test statistic.

Once the recurrence relation is stated, we are going to define the family of response-adaptive designs $\Re$, for which the algorithm is applicable. $\Re$ is the family of response-adaptive designs such that there exists a sequence of functions $\{h_n\}$ on $\mathbb{R}^3$ such that for each n

$$h_n(s,\, l) = P(\tilde{\delta}_n = 1|\, S_{n-1} = s,\ \Delta_{n-1} = l),$$

namely, the probability distribution of $\tilde{\delta}_n$ is completely determined by the values of S_n and Δ_n in the previous step.

The following proposition provides a recursion which will make computationally feasible to obtain the exact p-value for the S_n statistic when patients are randomized between treatments with a response-adaptive design of the family $\Re$.

Proposition 3.1.1. *Assume that n patients are allocated with a response-adaptive design in the family $\Re$. Then,*

$$\begin{aligned} P(S_n = s,\ \Delta_n = t) &= h_n(s - a_n,\, t-1)P(S_{n-1} = s - a_n,\ \Delta_{n-1} = t-1) + \\ &+ (1 - h_n(s + a_n,\, t+1))P(S_{n-1} = s + a_n,\ \Delta_{n-1} = t+1) \end{aligned}$$

with initial values for $a = 1,\, 0,\, -1$ and $b = 1,\, -1$

$$P(S_1 = a,\, \Delta_1 = b) = ((1 - a_1)(1 - |a|) + |a|a_1(1 + a * b)/2)P(\tilde{\delta}_n = b).$$

Proof. Observe that

$$\begin{aligned} P(S_n = s,\ \Delta_n = t) &= P(S_{n-1} + a_n\tilde{\delta}_n = s,\ \Delta_{n-1} + \tilde{\delta}_n = t) \\ &= P(S_{n-1} + a_n\tilde{\delta}_n = s,\ \Delta_{n-1} + \tilde{\delta}_n = t,\ \tilde{\delta}_n = 1) \\ &\quad + P(S_{n-1} + a_n\tilde{\delta}_n = s,\ \Delta_{n-1} + \tilde{\delta}_n = t,\ \tilde{\delta}_n = -1)\,. \end{aligned} \tag{3.4}$$

As

$$\begin{aligned} &P(S_{n-1} + a_n\tilde{\delta}_n = s,\ \Delta_{n-1} + \tilde{\delta}_n = t,\ \tilde{\delta}_n = 1) \\ &= P(\tilde{\delta}_n = 1|S_{n-1} = s - a_n, \Delta_{n-1} = t-1)P(S_{n-1} = s - a_n,\ \Delta_{n-1} = t-1) \end{aligned}$$

and

$$P(S_{n-1} + a_n\tilde{\delta}_n = s,\ \Delta_{n-1} + \tilde{\delta}_n = t,\ \tilde{\delta}_n = -1)$$
$$= \quad P(\tilde{\delta}_n = -1 | S_{n-1} = s + a_n, \Delta_{n-1} = t + 1) P(S_{n-1} = s + a_n,\ \Delta_{n-1} = t + 1),$$

then, the proposition follows. □

Remark 3.1.1. Proposition 6 in section 2.5 of [56] gives a formal argument, based on asymptotic tools, for choosing a permutation test as an alternative to the classical parametric test. Assuming the population model, when a response-adaptive design is used for allocations, and $S_{n,i}$ is the total number of successes with treatment i, $i = 1, 2$, then $(S_{n,1},\ S_{n,2},\ N_{n,1})$ are jointly sufficient for estimating p_1 and p_2; see p.193 in [63], where n is the sample size. So a good estimator of $p_1 - p_2$ is $S_{n,1}/N_{n,1} - S_{n,2}/N_{n,2}$, where $N_{n,2} = n - N_{n,1}$. Let $T_n = S_{n,1}/N_{n,1} - S_{n,2}/N_{n,2}$ be the difference between success proportions for each treatment when the n patients have been allocated with a response-adaptive design and the responses of the n patients, A_n, are known. Then,

$$\begin{aligned} P(T_n = t) &= \sum_{k=-n}^{n} P(T_n = t,\ \Delta_m = k) \\ &= \sum_{k=-n}^{n} P(S_n = r(k),\ \Delta_n = k) \end{aligned}$$

where $r(k) = (t(n^2 - k^2) + 2k\sum_{i=1}^{n} a_i)/(2n - k - \sum_{i=1}^{n} a_i)$. Now, using Proposition 1, any p-value for the distribution of T_n can be obtained.

The statistic T_n is also proposed and discussed in the context of treatment-adaptive designs in [9]. This proposal relies more on heuristic arguments than on theoretical ones.

3.2 The ℜ family of designs

As stated in the previous section, the family ℜ of response-adaptive designs collects those used to allocate n patients between two treatments and there exists a sequence of functions $\{h_n\}$ defined on $\mathbb{R}^2$ such that for each n,

$$h_n(s,\ l) = P(\tilde{\delta}_n = 1 |\ S_{n-1} = s,\ \Delta_{n-1} = l). \tag{3.5}$$

There are some useful relationships to link the probability of assigning treatment 1 with the statistics S_n and Δ_n. We know that

$$S_n = S_{n,1} - S_{n,2} \quad \text{and} \quad \sum_{i=1}^{n} a_i = S_{n,1} + S_{n,2}\,, \tag{3.6}$$

and also

$$\Delta_n = N_{n,1} - N_{n,2} \quad \text{and} \quad n = N_{n,1} + N_{n,2}\,. \tag{3.7}$$

These relations make it possible to write $S_{n,1}$, $S_{n,2}$, $N_{n,1}$ and $N_{n,2}$ in terms of S_n and Δ_n, the condition to be in the $\Re$ family is not stringent. Many designs in the literature belong to the family.

The Klein urn design has been presented and thoroughly studied under a population model in the previous chapter, see section 2.2. It works as follows. Initially, the urn contains $W_{0,1}$ balls of type 1, associated to treatment 1, and $2w - W_{0,1}$ balls of type 2, associated to treatment 2. $W_{0,1}$ could be any random variable with values in $E = \{0, 1, \cdots, w\}$, but usually we assume that $W_{0,1} = w$. When a patient arrives to the experiment a ball is extracted and receives the corresponding treatment. If the response of the patient is a success the ball is returned, otherwise, a ball of the other treatment is added. The total number of balls in the urn remains constant. The number of balls of type 1 in the urn is a stochastic process denoted by $\{W_{n,1}\}, n \in \mathbb{N}$.

Proposition 3.2.1. *For the Klein design Proposition 3.5 holds with*

$$h_n(r,\, l) = \frac{w + r - l}{2w} \quad \forall n\,.$$

Proof. From the dynamics of the Klein urn design, the probability of assigning treatment 1 is the number of type 1 of balls in the previous step,

$$P(\tilde{\delta}_n = 1) = \frac{W_{n-1}}{2w}\,.$$

The number of type 1 balls decreases in one unit in case of failure in treatment 1 and increases in one unit in case of failure in treatment 2. So, $W_n = w - (N_{n,1} - S_{n,1}) + (N_{n-2} - S_{n,2}) = w + S_n - \Delta_n$ and the result follows.

□

The next designs in the family are biased coin designs, such as DBCD designs, (1.9) and ERADE designs, (1.10). In these designs, the allocation rule uses the target allocation to assign patients. Under the population model, the target allocation depends on the parameters p_1 and p_2. Nevertheless, these quantities are unknown, so they are sequentially estimated, using the classical estimators $\hat{p}_{n1}$ and $\hat{p}_{n2}$, and these estimators are used to define the allocation rule. In the randomization model, the parametric approach disappears, but these estimators are still used in order to specify the allocation rule and reach a target allocation, although in fact they are not estimating a parameter. In a philosophical sense, we can substitute the concept of "estimation of the probability" by simply "proportion of successes". These are the proportions that are used in the following designs, which under the population model converge to the real value of the parameters with an initial value of 1/2 in both cases,

$$\hat{p}_{n1} = \frac{S_{n,1} + 1/2}{N_{n,1} + 1} \quad \text{and} \quad \hat{p}_{n2} = \frac{S_{n,2} + 1/2}{N_{n,2} + 1}\,. \tag{3.8}$$

Using the relations defined in (3.6) and (3.7), the estimators can be denoted in terms of S_n and Δ_n

$$\hat{p}_{n,1} = \frac{\sum_{i=1}^{n} a_i + S_n + 1}{n + \Delta_n + 2} \quad \text{and} \quad \hat{p}_{n,2} = \frac{\sum_{i=1}^{n} a_i - S_n + 1}{n - \Delta_n - 2}\,. \tag{3.9}$$

The *DBCD* designs, see (1.9), is a family of designs based on sequential estimation. For each n, the allocation rule is

$$P(\tilde{\delta}_{n+1} = 1|\mathcal{F}_n) = g_\alpha\left(\frac{N_{n,1}}{n},\, \hat{\rho}_n\right), \tag{3.10}$$

where $\hat{p}_{0,i} = 1/2$, with $i = 1,\, 2$ and $P(\delta_1 = 1) = 1/2$. The proposed probability function

is

$$g_\alpha(x,y)=\begin{cases}1, & x=0,\\ 0, & x=1,\\ \dfrac{y(y/x)^\alpha}{y(y/x)^\alpha+(1-y)((1-y)/(1-x))^\alpha}, & (x,y)\in(0,\,1)\times[0,\,1],\end{cases} \tag{3.11}$$

where the parameter x is referred to the proportion of treatment 1 allocations $N_{n,1}/n$ and the parameter y to $\hat{\rho}_n=\rho(\hat{p}_{n,1},\hat{p}_{n,2})$.

The Efficient randomized-adaptive designs (ERADE), see (1.10), have the following allocation rule,

$$P(\delta_{n+1}=1|\mathcal{F}_n)=\begin{cases}\alpha\hat{\rho}_n, & \text{if } \frac{N_{n,1}}{n}>\hat{\rho}_n;\\ \hat{\rho_n}, & \text{if } \frac{N_{n,1}}{n}=\hat{\rho}_n;\\ 1-\alpha(1-\hat{\rho}_n), & \text{if } \frac{N_{n,1}}{n}<\hat{\rho}_n.\end{cases} \tag{3.12}$$

In each step $\hat{\rho}_n=\rho(\hat{p}_{n,1},\hat{p}_{n,2})$ is calculated and it is compared to the actual proportion od allocations. If some treatment is underallocated, the probability of assigning this treatment is reduced, a kind of punishment, tuned with a parameter α.

In the last two designs the allocation rule is a function involving an approach of $\rho=\rho(p_1,p_2)$ by means of $\hat{\rho}_n=\rho(\hat{p}_{n,1},\hat{p}_{n,2})$ and the current proportion of treatment 1 allocations.

Proposition 3.2.2. *The DBCD designs and the ERADE designs are in the $\Re$ family.*

Proof. In both designs the allocation function is terms of $\hat{\rho}_n=\rho(\hat{p}_{n,1},\hat{p}_{n,2})$ and $N_{n,1}/n$. From (3.9) and the comments after it, we know that $\hat{\rho}_n$ can be written depending on S_n and Δ_n. Besides this, $N_{n,1}=\frac{n+\Delta_n}{2}$, so, only by function composition, the desired $h_n(s,l)$ is obtained. □

Remark 3.2.1. The previous proof is generic, it is valid for any target allocation and any value of the parameters of the design. It should be noted that it is not difficult to derive the expression of the $h_n(s,l)$ function in any case, but it is complicated and non

informative. For instante, for DBCD design with the most usual target allocation, RSIHR allocation and $\alpha = 2$, can be written as

$$h_{n+1}(r,\, l) = \frac{c_n^3(1-b_n)^2}{c_n^3(1-b_n)^2 + (1-c_n)^3 b_n^2}\,,$$

with $b_n = \dfrac{n+l}{2n}$ and $c_n = \dfrac{\sqrt{(\sum_{i=1}^n a_i) + r + 1}}{\sqrt{(\sum_{i=1}^n a_i) + r + 1} + \sqrt{\frac{n+l+1}{n-l+3}}\sqrt{(\sum_{i=1}^n a_i) - r + 1}}$.

Another outstanding design in the $\Re$ family is the play-the-winner rule, see section 1.1.3. In resume, when a patient arrives to a ball is extracted, the corresponding treatment is applied and the ball is returned with one ball of the same treatment in case of success and with one ball of the other treatment in case of failure.

Proposition 3.2.3. *The play-the-winner rule is in the $\Re$ family with*

$$h_{n+1}(r,\, l) = \frac{2w + n - l + 2r}{2(2w+n)}.$$

Proof. The probability of treatment 1 allocation in the $n+1$-th step is, as derived in (1.5),

$$P(\tilde{\delta}_{n+1} = 1 | \mathcal{F}_n) = \frac{w + S_{n,1} + (N_{n,2} - S_{n,2})}{2w+n}. \tag{3.13}$$

Using the relations in (3.6) and (3.7), the result follows. □

3.3 Simulation study

In the previous sections an algorithm to compute the exact distribution of the test statistic under RBI and hence the p-value was presented. Another common technique to compute the p-value is a Monte Carlo method. The main idea consists in fixing the size of the Monte Carlo estimator, K_c, and run the trial K_c times. In each run the test statistic is computed and compared with the observed value, already fixed. The proportion of times that the value of the statistic is greater than the observed one is the estimator of the p-value. One important concern about Monte Carlo method is the accuracy of the estimation. So, a comparative study is done between the exact p-value computed using the

algorithm presented in the previous section and Monte Carlo simulations of the p-values with two different values of K_c.

In Table 3.1 the results for two Monte Carlo simulations of p-values for different designs and the computation of the exact p-values are shown. The test statistic is the difference between successes in both treatments, $S_n = \sum a_i \tilde{\delta}_i = S_{n,1} - S_{n,2}$. The total number of patients is 50. $K_c = 2500$ and $K_c = 15000$ are Monte Carlo sample sizes, suggested in [57] as values that guarantee good levels of accuracy in the estimation. 1000 Monte Carlo runs are done, showing the mean (sd) of the 1000 runs. The null hypothesis is accepted if $|S_{50}| \geq 13$. In the exact computation, the exact distribution of the test statistic is computed, showing the corresponding probability tail.

Table 3.1: Monte Carlo estimation of p-value, mean (sd), and exact p-value. Test statistic S_n

	$K_c = 2500$	$K_c = 15000$	Exact
CR	0.0145 (0.0024)	0.0146 (0.0010)	0.0146
Efron	0.0013 (0.0007)	0.0013 (0.0003)	0.0013
Klein	0.0385 (0.0039)	0.0385 (0.0016)	0.0385
DBCD $\alpha = 0$	0.0791 (0.0055)	0.0791 (0.0022)	0.0793
DBCD $\alpha = 2$	0.0369 (0.0038)	0.0368 (0.0016)	0.0369
ERADE $\alpha = 0.5$	0.0318 (0.0035)	0.0317 (0.0014)	0.0317
ERADE $\alpha = 0.7$	0.0340 (0.0037)	0.0340 (0.0015)	0.0341
PTW	0.1326 (0.0110)	0.1327 (0.0027)	0.1326

If we consider the test statistic $T_n = S_{n,1}/N_{n,1} - S_{n,2}/N_{n,2}$, a similar computational study is done in Table 3.2. The set of responses is fixed and common for all the designs, and the observed experimental value of the test statistic is also common and fixed in $T = 0.3$, i.e. a difference of 0.3 between the percentage of successes in both treatments.

Table 3.2: Monte Carlo estimation of p-value, mean (sd), and exact value. Test statistic T_n

	$K_c = 2500$	$K_c = 15000$	Exact
CR	0.0340 (0.0034)	0.0341 (0.0015)	0.0341
Efron	0.0370 (0.0038)	0.0372 (0.0015)	0.0372
Klein	0.0287 (0.0034)	0.0286 (0.0014)	0.0286
DBCD $\alpha = 0$	0.0445 (0.0041)	0.0444 (0.0016)	0.0445
DBCD $\alpha = 2$	0.0383 (0.0040)	0.0383 (0.0016)	0.0382
ERADE $\alpha = 0.5$	0.0345 (0.0037)	0.0344 (0.0015)	0.0344
ERADE $\alpha = 0.7$	0.0384 (0.0039)	0.0383 (0.0016)	0.0384
PTW	0.0390 (0.0039)	0.0391 (0.0016)	0.0392

Observing Table 3.1 and Table 3.2, the first conclusion is that the mean estimations are almost equal to the exact ones in all the designs for both statistics. About the precision of the statistic, for the S_n statistic a dependence on the particular design is observed. Designs with large variability in the proportion of allocations are expected to have larger variability in the statistic. For instance, the case with higher variance in the Monte Carlo statistic is the PTW design, for which is well-known its high variability. Another remarkable issue is the uniformity of all the standard deviations in the T_n statistic.

In Table 3.3 a population model is assumed. Now, following [11], we have that under adaptive designs such that $N_{n,1} \to \infty$ and $N_{n,2} \to \infty$ as $n \to \infty$, the classical central limit theorem holds,

$$H(n, p_1, p_2) = \frac{\hat{p}_{n,1} - \hat{p}_{n,2} - (p_1 - p_2)}{\sqrt{\dfrac{p_1 q_1}{N_{n,1}} + \dfrac{p_2 q_2}{N_{n,2}}}} \to N(0, 1). \tag{3.14}$$

Then, for each pair of values of p_1 and p_2 and each allocation rule, $m = 1000$ executions of the corresponding clinical trial are made. Table 3.3 computes for several sample sizes n, $n = 30, 50, 100, 200, 500$, the proportion of times that $H(n, p_1, p_2) > 1.96$. That is, as we are in the setting of a population model, for a significance level of 0.05, the power of the Wald test is estimated.

As expected, for small values of n, $n \leq 50$, and small differences between p_1 and p_2, the power of the test is smaller than 0.5.

In Table 3.4 the same study is made but a population model is not assumed and now RBI is used with the S_n statistic. This is not an estimation of the power value, because this concept does not make sense in this setting, but it is the logical value to be obtained for comparative studies with power values, see [29], section 13.7, and [56]. The main idea is that for a randomization model does not exist an alternative hypothesis from which the data could be generated in order to do the simulation, as under a population model. So, the data, in our case the responses of the patients, are simulated using a parametric model, choosing two different success probabilities, and then the *power* is estimated as usual. Comparing the different designs, can be noted that the urn models have worse level of power in very small values of n. We observe that for small values of n and small differences of the success probabilities, *power* values are greater than in Table 3.3, except for the DCBD2 design.

Table 3.5 is as Table 3.4 but the test statistic T_n is chosen instead of S_n. Observe that both tables provide similar results except in the case $p_1 = p_2$ where the significance level is correctly reached for all the designs even for small values of n. This is quite remarkable because mitigates the conservativeness of the S_n statistic. We also observe that it takes the increment of the power for small values of n which was observed in Table 3.4 but for large values of n the behavior is as its population statistical test counterpart, which is used in Table 3.3.

Table 3.3: Proportion of rejections with the Wald test: asymptotic estimations

		Sample size			
p_1, p_2	D	10	30	50	100
0.9, 0.3	CR	0.503	0.977	0.998	1.000
	Efron	0.603	0.979	0.999	1.000
	PTW	0.393	0.904	0.987	1.000
	Klein	0.470	0.956	1.000	1.000
	DBCD2	0.588	0.978	0.999	1.000
	ERADE7	0.494	0.971	0.999	1.000
0.9, 0.5	CR	0.170	0.710	0.925	0.995
	Efron	0.150	0.742	0.932	0.997
	PTW	0.137	0.602	0.830	0.978
	Klein	0.159	0.689	0.881	0.992
	DBCD2	0.157	0.715	0.925	0.997
	ERADE7	0.149	0.719	0.892	0.995
0.9, 0.7	CR	0.076	0.252	0.421	0.701
	Efron	0.074	0.232	0.412	0.730
	PTW	0.048	0.228	0.394	0.634
	Klein	0.071	0.251	0.410	0.689
	DBCD2	0.083	0.248	0.418	0.705
	ERADE7	0.067	0.246	0.416	0.698
0.7, 0.3	CR	0.124	0.587	0.823	0.978
	Efron	0.158	0.631	0.865	0.988
	PTW	0.098	0.582	0.842	0.992
	Klein	0.119	0.614	0.838	0.989
	DBCD2	0.182	0.626	0.855	0.993
	ERADE7	0.128	0.618	0.832	0.982
0.7, 0.5	CR	0.046	0.179	0.312	0.525
	Efron	0.038	0.130	0.294	0.527
	PTW	0.035	0.174	0.261	0.514
	Klein	0.046	0.174	0.296	0.520
	DBCD2	0.046	0.190	0.312	0.532
	ERADE7	0.051	0.169	0.273	0.513
0.7, 0.7	CR	0.019	0.042	0.040	0.041
	Efron	0.013	0.040	0.046	0.046
	PTW	0.015	0.043	0.052	0.043
	Klein	0.017	0.046	0.050	0.046
	DBCD2	0.012	0.037	0.043	0.039
	ERADE7	0.017	0.043	0.036	0.040

Table 3.4: Proportion of rejections with the randomization test (exact p-values) with test statistic S_n.

		Sample size			
p_1, p_2	D	10	30	50	100
0.9, 0.3	CR	0.382	0.705	0.861	0.994
	Efron	0.656	0.963	0.998	1.000
	PTW	0.028	0.611	0.870	0.999
	Klein	0.169	0.843	0.984	1.000
	DBCD2	0.315	0.659	0.904	0.999
	ERADE7	0.606	0.952	0.997	1.000
0.9, 0.5	CR	0.247	0.330	0.475	0.713
	Efron	0.399	0.736	0.912	0.997
	PTW	0.025	0.267	0.483	0.808
	Klein	0.072	0.441	0.756	0.981
	DBCD2	0.170	0.327	0.529	0.880
	ERADE7	0.314	0.689	0.890	0.994
0.9, 0.7	CR	0.136	0.151	0.175	0.241
	Efron	0.230	0.308	0.452	0.739
	PTW	0.025	0.085	0.132	0.243
	Klein	0.023	0.122	0.271	0.605
	DBCD2	0.137	0.097	0.141	0.293
	ERADE7	0.182	0.308	0.460	0.728
0.7, 0.3	CR	0.324	0.468	0.602	0.854
	Efron	0.447	0.722	0.879	0.989
	PTW	0.005	0.307	0.575	0.916
	Klein	0.069	0.419	0.732	0.969
	DBCD2	0.285	0.284	0.488	0.868
	ERADE7	0.400	0.630	0.815	0.981
0.7, 0.5	CR	0.184	0.159	0.182	0.277
	Efron	0.224	0.271	0.355	0.576
	PTW	0.006	0.086	0.146	0.302
	Klein	0.018	0.121	0.220	0.458
	DBCD2	0.115	0.072	0.096	0.204
	ERADE7	0.199	0.242	0.326	0.552
0.7, 0.7	CR	0.121	0.072	0.063	0.064
	Efron	0.200	0.112	0.120	0.081
	PTW	0.010	0.036	0.034	0.044
	Klein	0.007	0.017	0.020	0.015
	DBCD2	0.099	0.016	0.010	0.009
	ERADE7	0.134	0.087	0.072	0.071

Table 3.5: Proportion of rejections with the randomization test (Monte Carlo estimations of the p-values) and test statistic T.

		Sample size			
p_1, p_2	D	10	30	50	100
0.9, 0.3	CR	0.552	0.960	0.999	1.000
	Efron	0.657	0.969	1.000	1.000
	PTW	0.334	0.919	0.996	1.000
	Klein	0.430	0.948	0.995	1.000
	DBCD2	0.423	0.955	0.998	1.000
	ERADE7	0.349	0.938	0.996	1.000
0.9, 0.5	CR	0.325	0.699	0.895	0.998
	Efron	0.385	0.765	0.927	0.999
	PTW	0.164	0.666	0.870	0.991
	Klein	0.196	0.676	0.863	0.994
	DBCD2	0.263	0.691	0.896	0.996
	ERADE7	0.152	0.646	0.865	0.995
0.9, 0.7	CR	0.157	0.278	0.439	0.754
	Efron	0.202	0.347	0.487	0.750
	PTW	0.066	0.216	0.377	0.670
	Klein	0.053	0.217	0.395	0.696
	DBCD2	0.160	0.262	0.440	0.717
	ERADE7	0.060	0.229	0.390	0.717
0.7, 0.3	CR	0.239	0.614	0.822	0.981
	Efron	0.350	0.677	0.860	0.986
	PTW	0.156	0.539	0.775	0.986
	Klein	0.163	0.567	0.804	0.978
	DBCD2	0.298	0.559	0.802	0.982
	ERADE7	0.146	0.544	0.808	0.980
0.7, 0.5	CR	0.136	0.227	0.328	0.544
	Efron	0.168	0.226	0.328	0.564
	PTW	0.075	0.171	0.273	0.493
	Klein	0.063	0.181	0.266	0.520
	DBCD2	0.118	0.182	0.263	0.494
	ERADE7	0.043	0.157	0.265	0.518
0.7, 0.7	CR	0.062	0.060	0.043	0.051
	Efron	0.092	0.073	0.069	0.073
	PTW	0.030	0.048	0.038	0.046
	Klein	0.028	0.042	0.042	0.040
	DBCD2	0.084	0.053	0.045	0.046
	ERADE7	0.030	0.026	0.044	0.046

3.4 Randomization based inference for the DTL design

Although the $\Re$ family includes most of the response-adaptive designs considered in the literature, the designs based on immigrated urn models, see [79], do not belong to it. The DTL design, see (1.6), is one of them. When a treatment ball is extracted, the patient is allocated to this treatment, as usual. But when an immigration ball is extracted, no treatment is applied and a change in the urn composition is done, so there is a break up between ball extractions and treatment allocations. Due to this break up, these type of designs cannot be included in the $\Re$ family. The probability of allocation is no longer a function of S_n and Δ_n, because the number of immigration balls drawn until the present allocation is also a required information to determine the probability distribution of the allocation rule. Furthermore, the number of consecutive immigration ball extractions can be any non-negative number, so the probability of treatment 1 allocation only can be expressed by an infinite sum, not in a closed function form.

In [79], theoretical properties for a wide range of urn models have been presented under an unified approach called immigrated urn models. In addition to this, applications of immigrated urn models in clinical trials are also studied under the assumption of a population model. We assume the randomization model, with a fixed set of responses $A_n = \{Y_1 = a_1, Y_2 = a_2, \ldots, Y_n = a_n\}$, so the results in [79] are not applicable in our case. For this reason, a particular study of RBI for these designs is of interest. We will focus on the DTL design, which behavior was presented in section 1.1.3. The number of type 1 balls in the urn after the mth replacement and the total number of balls are denoted, respectively, by $W_{m,1}$ and T_m. So $W_{0,1} = w$ and $T_0 = 2w + 1$. Observe that, due to the presence of a type 0 ball, the number of replacements, m, is not necessarily equal to the number of patients, n, that have participated in the experiment, being $n \leq m$.

The theoretical treatment of the model is complicated, due to the duality between urn extractions and treatment allocations. In order to express the model mathematically we need to introduce some notation. For each extraction m, $m = 1, 2, \cdots$, φ_m is a random

variable which equals 1 when a ball of type 1 is drawn, -1 when the ball is of type 2 and 0 when the ball is of type 0. Let $\Delta_m = \sum_{i=1}^{m} \varphi_i$ be the unbalance between the number of patients allocated to each treatment after the m-th replacement. If $N_{m,1}$ and $N_{m,2}$ are, as usual, the number of patients allocated to treatment 1 and treatment 2 up to and including the mth replacement, then $\Delta_m = \sum_{i=1}^{m} \varphi_i = N_{m,1} - N_{m,2}$. If we denote as $N_{m,0}$ the number of type 0 balls drawn up to the mth replacement, we can express this quantity as $N_{m,0} = m - N_{m,1} - N_{m,2}$.

Finally, let $\mathcal{F}_m = \sigma(\mathcal{G}_m, \mathcal{Z}_m)$ denote the natural sigma algebra generated by all the previous allocations and responses up to the m-th replacement. Then,

$$P(\varphi_{m+1} = 1|\mathcal{F}_m) = \frac{W_{m,1}}{T_m}, \quad P(\varphi_{m+1} = 0|\mathcal{F}_m) = \frac{1}{T_m}. \tag{3.15}$$

As the number of extractions does not correspond with the number of treatment allocations, additional notation should be included. Let $\{\tau_m\}_{m\geq 0}$ be a sequence of stopping times which represents, for each m, the number of patients allocated up to the mth replacement. Observe that taking $\tau_0 = 0$ then, for $m \geq 1$, $\tau_m = \sum_{i=1}^{m} \varphi_i^2$, because φ_i^2 is equal to 1 if any treatment has been applied and equal to 0 if an immigration ball has been extracted and then, there has not been any treatment allocation. It also holds that $\tau_m = \sum_{i=1}^{m} \varphi_i^2 = N_{m,1} + N_{m,2}$. Observe that using the expressions from above, $N_{m,1}$ and $N_{m,2}$ can be expressed as $N_{m,1} = \sum_{i=1}^{m}(\varphi_i + \varphi_i^2)/2$ and $N_{m,2} = \sum_{i=1}^{m}(\varphi_i^2 - \varphi_i)/2$.

Let S_n be the difference between the number of successes in the two treatments once n patients have responded under the DTL rule. Observe that a type 0 ball can be extracted in each step and this implies that the number of extractions of type 0 balls between two allocations is an unbounded random variable. It will be helpful to define S_m^*, the difference between the number of successes up to and including the mth extraction, so consider $S_m^* = \sum_{i=1}^{m} a_{\tau_i}\varphi_i$.

As in the previous sections, S_n is chosen as the randomization test statistic for studying the equality of treatments. We are interested in computing the exact distribution of S_n. Observe that the event $\{S_n = s\}$ can be expressed as the a union of disjoint events indexed

by $m \geq n$, and then,

$$P(S_n = s) = \sum_{m=n}^{\infty} P(S_m^* = s,\ \tau_m = n,\ \varphi_m \neq 0). \tag{3.16}$$

Note that this decomposition is significantly different from the decomposition in (3.2). Now, the probability is an infinite sum, because of the possibility of extracting any number of immigrations balls without applying any treatment allocation. The following technical proposition will be crucial for computing the exact p-values of the randomization test based on the S_n statistic for the DTL design. The algorithm is inspired by Proposition 3.1.1, adapted to the new urn with three different type of balls.

Proposition 3.4.1. *Assume that the DTL design is applied as allocation rule. Consider for each replacement $m \geq 1$ the random variables φ_m, Δ_m and τ_m defined as before. Then,*

$$P(S_m^* = s,\ \tau_m = n,\ \varphi_m \neq 0),\ -n \leq s \leq n, \tag{3.17}$$

can be calculated with a recursion formula with initial values

$$\begin{aligned} P(S_1^* = 0,\ \tau_1 = 0,\ \Delta_1 = 0) &= \frac{1}{2w+1}, \\ P(S_1^* = 0,\ \tau_1 = 1,\ \Delta_1 = -1) = P(S_1^* = 0,\ \tau_1 = 1,\ \Delta_1 = 1) &= (1-a_1)\frac{w}{2w+1}, \\ P(S_1^* = -1,\ \tau_1 = 1,\ \Delta_1 = -1) = P(S_1^* = 1,\ \tau_1 = 1,\ \Delta_1 = 1) &= a_1\frac{w}{2w+1}. \end{aligned}$$

Proof. The probability in (3.17) can be decomposed as

$$P(S_m^* = s,\ \tau_m = n,\ \varphi_m = r) = \sum_{i=-n}^{n} P(S_m^* = s,\ \tau_m = n,\ \Delta_m = i,\ \varphi_m = r).$$

Observe that for each extraction $m \geq 1$ we have

$$\begin{aligned} &P(S_m^* = s,\ \tau_m = k,\ \Delta_m = i,\ \varphi_m = r) = \\ &P(\varphi_m = r | S_{m-1}^* = s - a_k r,\ \tau_{m-1} = k - r^2,\ \Delta_{m-1} = i - r) \\ &\times P(S_{m-1}^* = s - a_k r,\ \tau_{m-1} = k - r^2,\ \Delta_{m-1} = i - r). \end{aligned} \tag{3.18}$$

On the other hand, in order to have a solvable recurrence, we need to express the number of total balls and type 1 balls as a function of the elements in the conditional part of

(3.18). Considering that after each immigration ball two other balls are added to the urn and after a success in a treatment one ball is dropped, we can express the total number of balls as

$$T_m = 2w + 1 + 2N_{m,0} - \left(\tau_m - \sum_{j=1}^{\tau_m} a_j \right) = 2w + 1 + 2m - 3\tau_m + \sum_{j=1}^{\tau_m} a_j \, .$$

A similar calculation for the number of type 1 balls leads to

$$W_{m,1} = \frac{T_m - 1 + S_m^* - \Delta_m}{2};$$

so, for $r = 1$ and $r = 0$ we can explicitly express

$$P(\varphi_m = 1 | S_{m-1}^* = l_1,\ \tau_{m-1} = l_2,\ \Delta_{m-1} = l_3) = \frac{2w+1+2m-3l_2+\sum_{j=1}^{l_2} a_j - 1 + l_1 - l_3}{2\left(2w+1+2m-3l_2+\sum_{j=1}^{l_2} a_j\right)},$$

$$P(\varphi_m = 0 | S_{m-1}^* = l_1,\ \tau_{m-1} = l_2,\ \Delta_{m-1} = l_3) = \frac{1}{2w+1+2m-3l_2+\sum_{j=1}^{l_2} a_j}.$$

A closed expression for all the values of r can be obtained,

$$P(\varphi_m = r | S_{m-1}^* = l_1,\ \tau_{m-1} = l_2,\ \Delta_{m-1} = l_3) = \frac{r^2(l_1 - l_3)) + r(T_{m-1} - 1) + 2}{2T_{m-1}} \, . \quad (3.19)$$

Finally, a recursion is obtained by plugging (3.19) into (3.18). □

When randomization is made with the DTL design, Proposition 3.4.1 and equation (3.16) allow us to obtain exact calculations of p-values when RBI is used with the test statistic S_n.

Remark 3.4.1. As can be seen in (3.16), the probability $P(S_n = s)$ is an infinite sume of elements. With the recursive algorithm each summand can be computed, but to compute the whole summation an approximation level must be fixed. This level is going to be an error in the exact calculation, however, this error could be as small as we want.

Remark 3.4.2. As happened in Remark 3.1.1 with the $\Re$ family of designs, other test statistics distributions can be computed exactly when the DTL rule is applied. A good

estimator of $p_1 - p_2$ is $S_{n,1}/N_{n,1} - S_{n,2}/N_{n,2}$. Let T_n be the difference between success proportions for each treatment when the n patients have been allocated with the DTL rule. Reasoning as in (3.16), it follows that if T_m^* is the difference of success proportions up to and including the mth ball extraction, then,

$$\begin{aligned} P(T_n = t) &= \sum_{m=n}^{\infty} \sum_{k=0}^{n} P(T_m^* = t,\ \Delta_m = 2k - n,\ \tau_m = n,\ \varphi_m \neq 0) \\ &= \sum_{m=n}^{\infty} \sum_{k=0}^{n} P(S_m^* = r(k),\ \Delta_m = 2k - n,\ \tau_m = n,\ \varphi_m \neq 0) \end{aligned}$$

where $r(k) = (2t * k * (n-k) - (n-2k)\sum_{i=1}^{n} a_i)/n$. Now, using Proposition 3.4.1, any p-value for the distribution of T_n can be obtained.

In Table 3.6 we replicate the procedure of Table 3.1 for the DTL design, using the recurrence algorithm for this design. Comparing the computed p-values with the results for the Klein urn design, there are significant differences, so even both designs show a similar asymptotic behavior, this result suggests some differences between the behavior of both designs in small or moderate sample sizes.

Table 3.6: Monte Carlo estimation of p-value, mean (sd), and exact p-value for the DTL design. 50 patients. Test statistic S_n

	$K_c = 2500$	$K_c = 15000$	Exact
CR	0.0145 (0.0024)	0.0146 (0.0010)	0.0146
Klein	0.0385 (0.0039)	0.0385 (0.0016)	0.0385
DTL	0.0596 (0.0048)	0.0595 (0.0020)	0.0595

Chapter 4

Asymptotic randomization based inference for the Klein urn design

Randomization based inference has been shown as an adequate alternative to population based inference when the assumption of the population model is not appropriate. However, some of the principal drawbacks of this kind of inference is due to the enumeration and the computation of the probability of all the possible data permutations. This could be computationally unfeasible. In order to avoid this problem, one option is the development of algorithms to shorten the calculations, as the one presented in Chapter 3. Other alternative is to obtain the asymptotic distribution of the test statistic. The computation of the asymptotic distribution of the randomization test under the Klein urn design has two main problems. One the one hand, due to the particular dependence structure of the allocations, standard central limit theorems are not valid and a specific calculation needs to be done. On the other hand, when the randomization model is assumed, the responses are considered fixed, so the dynamics of the design changes and the stochastic properties have to be studied under the new paradigm. In section 1 the Klein urn under the randomization model is presented and studied, partially replicating the results of Chapter 2. In section 2, the asymptotic distribution of the randomization test statistic is computed under certain

conditions, also analyzing the adequacy of these conditions.

The contents of this chapter have been collected in [25].

4.1 Klein urn design under randomization model

Under a population model, the response of the nth patient is $Y_n = \delta_n Z_{n,1} + (1-\delta_n)Z_{n,2}$, where, for $i = 1, 2$, $Z_{n,i}$ are independent and identically distributed random variables with Bernoulli distribution of parameter p_i. The fact of having these random responses has a strong influence in the behavior of the urn and therefore in the properties of the design in general. For example, the evolution of the number of balls of type 1 in the urn, $\{W_{n,1}\}, n \in \mathbb{N}$, the process which mainly determines the behavior of the design, is a homogeneous Markov chain with states $E = \{0, 1, \ldots, 2w\}$ and transition probabilities

$$p_{i,j} = \begin{cases} q_2(1-\frac{i}{2w}), & j = i+1, \\ p_1\frac{i}{2w} + p_2(1-\frac{i}{2w}), & j = i, \\ q_1\frac{i}{2w}, & j = i-1, \\ 0, & \text{otherwise.} \end{cases} \qquad i, j = 0, \ldots, 2w.$$

When the response of a patient is a success, the urn composition remains unchanged, when it is a failure, we do not return the ball and a ball of the other type is added. This dual behavior is due to the random binary responses, and the Markov chain $W_{n,1}$ has a homogeneous representation. Under the randomization model, the responses are deterministic so both behaviors must be distinguished. In the sequel, the set of responses $A_n = \{Y_1 = a_1, Y_2 = a_2, \cdots, Y_n = a_n\}$ is considered fixed and all the random variables are conditioned to this set of responses. Analyzing the new urn model, when the (deterministic) response after any treatment application is a success, we do not change the composition of the urn. In case of a (deterministic) failure, the number of balls of the applied treatment decreases one unit and the number of balls of the other type increases one unit. The behavior of the urn in case of failure is the same as the Ehrenfest urn model. Under this new paradigm, the number of type 1 balls in the urn is no longer a

homogeneous Markov chain. The markovian property is not lost, but due to the duality in the behavior of $\{W_{n,1}\}$, the chain is now a non-homogeneous Markov chain. Besides this, the expectation and the variance of the variables of the process must be recalculated under the new paradigm.

Proposition 4.1.1. *Assume that the Klein design has been applied to a set of n patients and the responses are A_n. The initial number of balls of type 1 is a random variable $W_{0,1}$ such that $E[W_{0,1}] = w$. Then,*

a) The process $\{W_{n,1}\}$ is a non-homogeneous Markov Chain with state space $E = \{0, 1, \ldots, 2w\}$ such that, for each n, the transition probability matrix is $P_n = a_n I + (1 - a_n)P$. Where I is the identity matrix of order $2w + 1$ and P is the transition matrix of the classical Ehrenfest Urn model

$$P(i,j) = \begin{cases} 1 - \dfrac{i}{2w}, & j = i + 1 \\ \dfrac{i}{2w}, & j = i - 1 \qquad i, j \in E. \\ 0, & \textit{otherwise.} \end{cases}$$

b) Let $f_n = \sum_{i=1}^{n}(1 - a_i)$ be the number of failures up to n, then, for each n,

$$E[W_{n,1}] = w \text{ and } Var[W_{n,1}] = \frac{w}{2}\left(1 - \left(1 - \frac{2}{w}\right)^{f_n}\left(1 - \frac{2}{w}Var[W_{0,1}]\right)\right).$$

Proof. For each n we have

$$W_{n,1} = W_{n-1,1} - (1 - a_n)\tilde{\delta}_n \tag{4.1}$$

To prove a), observe that when the response of the i-th patient is successful, the chain does not change, $W_{n,1} = W_{n-1,1}$. Otherwise, it jumps to the neighbor states as in the classical Ehrenfest process. So, the transition matrix is the combination of both transitions matrices, depending on the deterministic responses.

Taking expectations in (4.1), we have that

$$\begin{aligned} E[W_{n,1} - w] &= E[W_{n-1,1} - w] - (1 - a_n)E\left[E[\tilde{\delta}_n | W_{n-1,1}]\right] \\ &= \left(1 - \frac{1 - a_n}{w}\right) E[W_{n-1,1} - w] \\ &= \prod_{i=1}^{n} \left(1 - \frac{1 - a_i}{w}\right) E[W_{0,1} - w]\,, \end{aligned}$$

as $E[W_{0,1}] = w$, then $E[W_{n,1}] = w$. Also from (4.1) we have

$$\begin{aligned} E\left[(W_{n,1} - w)^2\right] &= E\left[(W_{n-1,1} - w)^2\right] + (1 - a_n) - 2(1 - a_n)E\left[(W_{n-1,1} - w)\,\tilde{\delta}_n\right] \\ &= \left(1 - 2\frac{1 - a_n}{w}\right) E[(W_{n-1,1} - w)^2] + (1 - a_n)\,, \end{aligned}$$

so that, if $a_n = 1$ we have that $Var[W_{n,1}] = Var[W_{n-1,1}]$, otherwise,

$$\begin{aligned} Var[W_{n,1}] &= \left(1 - \frac{2}{w}\right) Var\left[W_{n-1,1}\right] + 1 \\ &= \sum_{i=0}^{f_n - 1} \left(1 - \frac{2}{w}\right)^i + \left(1 - \frac{2}{w}\right)^{f_n} Var[W_{0,1}]\,. \end{aligned}$$

Then, $Var[W_{n,1}] = \dfrac{w}{2}\left(1 - \left(1 - \dfrac{2}{w}\right)^{f_n}\left(1 - \dfrac{2}{w} Var[W_{0,1}]\right)\right)$ as stated in b).

□

Remark 4.1.1. Note how in the number of type 1 balls process, $\{W_{n,1}\}$, the random variables are centered for each n considering the urn state space $[0, 2w]$, when the responses are fixed. It can be understood if we think that the number of balls in the urn is skewed when a treatment performs better than another treatment, i.e. when the success probability of one treatment is greater than the other one, if we were under the population model. In our case, under the randomization model, the treatments are considered equivalent and the responses are fixed, so the concept of probability of success disappears, but we can think that the urn performs as the success probabilities were equal. For this reason, the expected number of balls is w for each n.

Remark 4.1.2. Observing the expression of the variance of $W_{n,1}$ in Proposition 4.1.1, some conclusion can be obtained. If the number of failures converges to infinity, $f_n \to \infty$,

then $Var[W_{n,1}] \to \frac{w}{2}$ and this convergence is monotonous, since $\left(1-\frac{2}{w}\right)^{f_n}$ is monotonous. Besides this,

if $Var[W_{0,1}] > \frac{w}{2}$ then, $\{Var[W_{n,1}]\}$ is strictly decreasing,

if $Var[W_{0,1}] = \frac{w}{2}$ then, $\{Var[W_{n,1}]\}$ is equal to $\frac{w}{2}$ for each n,

if $Var[W_{0,1}] < \frac{w}{2}$ then, $\{Var[W_{n,1}]\}$ is strictly increasing.

In Chapter 2, during the study of the properties of the Klein urn design under a population model, we have emphasized the importance of the fact that the number of balls of type 1 is a homogeneous Markov chain and the importance of this property in the study of almost all the processes of the design. Under the randomization model, $\{W_{n,1}\}$ is non-homogeneous Markov chain, so an important change in the treatment's allocation process $\{\delta_n\}$ is expected. In the next proposition new properties of the process $\{\delta_n\}$ are studied.

Proposition 4.1.2. *Assume that the Klein design has been applied to a set of n patients and the responses are A_n. Then,*

a) For each n

$$E[\delta_n] = \frac{1}{2} \text{ and } Var[\delta_n] = \frac{1}{4}$$

and

$$E[\tilde{\delta}_n] = 0 \text{ and } Var[\tilde{\delta}_n] = 1$$

b) The components of the covariance matrix of the vector $\boldsymbol{\delta}_n = \{\delta_1, ..., \delta_n\}$, are

$$\Sigma_{\boldsymbol{\delta}_n}(i,\, j) = \begin{cases} Var[\delta_i], & i = j \\ \dfrac{1}{4w^2} Var[W_{i-1,1}] - \dfrac{(1-a_i)}{4w}, & j = i+1; \\ \left(1-\dfrac{1}{w}\right)^{f_j - f_i} Cov(\delta_i, \delta_{i+1}), & j > i+1 \end{cases}$$

c) The covariance matrix of the sequence of martingale differences $\{\delta_i - E[\delta_i|\mathcal{F}_{i-1}]\}$, $i = 1, \ldots, n$, is a diagonal matrix with

$$\Sigma_{\boldsymbol{\delta}_n - E[\boldsymbol{\delta}_n|\mathcal{F}_{n-1}]}(i,\, i) = \frac{1}{4}\left(1 - \frac{1}{w^2} Var[W_{i-1,1}]\right).$$

Proof. From Proposition 4.1.2 part b), $E[W_{n-1,1}] = w$, so,

$$E[\delta_n] = E[W_{n-1,1}/(2w)] = \frac{1}{2}\,.$$

For the variance,

$$Var[\delta_n] = E\left[\left(\delta_n - \frac{1}{2}\right)^2\right] = \frac{1}{4}\,,$$

and a) follows.

For b) observe that

$$\begin{aligned}
Cov(\delta_i,\, \delta_{i+1}) &= E[\delta_i E[\delta_{i+1}|W_{i,1}] - E[\delta_i]E[\delta_{i+1}] \\
&= \frac{1}{2w} E[\delta_i W_{i,1}] - \frac{1}{4} \\
&= \frac{1}{2w} E[\delta_i (W_{i-1,1} - (1-a_i)(2\delta_i - 1))] - \frac{1}{4} \\
&= \frac{1}{2w} (E[\delta_i W_{i-1,1}] - (1-a_i)E[\delta_i]) - \frac{1}{4} \\
&= \frac{1}{4w^2} (E[(W_{i-1,1})^2] - (1-a_i)w) - \frac{1}{4} \\
&= \frac{1}{4w^2} Var[W_{i-1,1}] - \frac{1-a_i}{4w}\,,
\end{aligned}$$

and also, for $j > i+1$,

$$\begin{aligned}
Cov(\delta_i,\, \delta_j) &= \frac{1}{2w} E[\delta_i W_{j-1,1}] - \frac{1}{4} \\
&= \frac{1}{2w} E\left[\delta_i (W_{j-2,1} - (1-a_{j-1})(2\delta_{j-1} - 1))\right] - \frac{1}{4} \\
&= \frac{1}{2w}\left(E[\delta_i W_{j-2,1}] - (1-a_{j-1})\left(2E[\delta_i \delta_{j-1}] - \frac{1}{2}\right)\right) - \frac{1}{4} \\
&= E[\delta_i \delta_{j-1}] - \frac{1-a_{j-1}}{w}\left(E[\delta_i \delta_{j-1}] - \frac{1}{4}\right) - \frac{1}{4} \\
&= \left(1 - \frac{1-a_{j-1}}{w}\right) Cov(\delta_i,\, \delta_{j-1})\,,
\end{aligned}$$

and c) follows.

Observe that for any $j > i$, with $i, j = 1, \ldots, n$ we have that

$$\begin{aligned} Cov(\delta_i - E[\delta_i|\mathcal{F}_{i-1}], \delta_j - E[\delta_j|\mathcal{F}_{j-1}]) &= Cov(\delta_i, \delta_j) - Cov(\delta_i, E[\delta_j|\mathcal{F}_{j-1}) \\ &- Cov(E[\delta_i|\mathcal{F}_{i-1}], \delta_j) \\ &+ Cov(E[\delta_i|\mathcal{F}_{i-1}], E[\delta_j|\mathcal{F}_{j-1}]) \,. \end{aligned}$$

From Lemma 2.2.1, and noting that δ_i and $E[\delta_i|\mathcal{F}_{i-1}]$ are $\mathcal{F}_{j-1}$-measurable,

$$Cov(\delta_i, \delta_j) = Cov(\delta_i, E[\delta_j|\mathcal{F}_{j-1}) \text{ and } Cov(E[\delta_i|\mathcal{F}_{i-1}], \delta_j) = Cov(E[\delta_i|\mathcal{F}_{i-1}], E[\delta_j|\mathcal{F}_{j-1}]),$$

then, the set of variables $\{\delta_i - E[\delta_i|W_{i-1}]\}_{i=1,\ldots,n}$ are incorrelated. Besides this,

$$\begin{aligned} Var(\delta_i - E[\delta_i|W_{i-1,1}]) &= E[\delta_i] - E\left[\frac{(W_{i-1,1})^2}{4w^2}\right] \\ &= \frac{1}{2} - \frac{Var[W_{i-1,1}] + w^2}{4w^2} \\ &= \frac{1}{4} - \frac{Var[W_{i-1,1}]}{4w^2}, \end{aligned}$$

so, the proposition is proved. □

Remark 4.1.3. Note that $\tilde{\delta}_n = 2\delta_n - 1$ and $(\tilde{\delta}_n - E[\tilde{\delta}_n|\mathcal{F}_{n-1}]) = 2\,(\delta_n - E[\delta_n|\mathcal{F}_{n-1}])$, so, the covariance matrices follow the same relation, $\Sigma_{\boldsymbol{\delta}_n} = 4\Sigma_{\tilde{\boldsymbol{\delta}}_n}$ and $\Sigma_{\boldsymbol{\delta}_n - E[\boldsymbol{\delta}_n|\mathcal{F}_{n-1}]} = 4\Sigma_{\tilde{\boldsymbol{\delta}}_n - E[\tilde{\boldsymbol{\delta}}_n|\mathcal{F}_{n-1}]}$.

In order to continue the study of the properties of the design, the presence of the non-homogeneous Markov chain $\{W_{n,1}\}$ makes it more difficult, specially in asymptotic theory. The principal concern about the asymptotic behavior of a Markov chain is the existence of an invariant or long run distribution. This property is called ergodicity, and it can be thought as a kind of stabilization of the chain. A chaotic behavior of the transition matrix would make the asymptotic study unfeasible. If some kind of stabilization could be guaranteed, the asymptotic study would become reasonable. In [38], Chapter V, a study of different types of ergodic properties for non-homogeneous Markov chains is done. Almost all the results are based on the Dobrushin's ergodic coefficient.

In the Klein urn design under the randomization model, as it has been seen in Proposition 4.1.1, the process $\{W_{n,1}\}$ gets stuck or moves randomly as in the classical Ehrenfest model, so it is a deterministic combination of two different and well known chains. Using this property, the ergodicity of $\{W_{n,1}\}$ is proven in the next lemma.

Lemma 4.1.1. *Assume the conditions of Proposition 4.1.2 and that* $f_n = \sum(1 - a_i) \uparrow \infty$ *when* $n \to \infty$. *Then, the non-homogeneous Markov-Chain* $\{W_{n,1}\}$ *is strongly ergodic.*

Proof. From Proposition 4.1.2 we have that $\{W_{n,1}\}$ is a non-homogeneous Markov Chain with transition probability $P_n = a_n I + (1-a_n)P$. In case of a success the transition matrix is the identity and in case of failure it is the Ehrenfest urn model transition matrix. The n-step transition matrix has an easy form, $P^{(k,k+n)} = \prod_{i=k}^{k+n} P_i = P^{f_{k+n}-f_k}$, depending only in the number of failures. As $n \to \infty$, $f_n \to \infty$ and then

$$P^{(k,k+n)} = \prod_{i=k}^{k+n} P_i = P^{f_{k+n}-f_k} \to \mathbf{1}^t \boldsymbol{\pi}\,.$$

where $\mathbf{1}^t$ is a column vector of ones and $\boldsymbol{\pi}$ is the stationary distribution of the classical Ehrenfest urn model, which is a binomial distribution with parameters $2w$ and $1/2$, see, for instance, Proposition 2.1.2. The n-step transition matrix converges to a constant matrix when n goes to infinity, so then, from Theorem V.4.1 in [38], the result follows. □

Strong laws and central limit theorems for non-homogeneous Markov Chains remain still as a research topic in Probability Theory. Some advances can be found in the literature to obtain a strong law of large numbers for $\{W_{n,1}\}$. Here a modification of a result in [44] is presented.

Proposition 4.1.3. *Assume the conditions of Proposition 4.1.2. Let* $\{g_n\}$, $n \geq 0$, *be a sequence of real functions defined on the state space* E. *If* $|g_n(W_{n,1})| \leq M$, $n \geq 0$ *and* $f_n \uparrow \infty$ *when* $n \to \infty$, *then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (g_i(W_{i,1}) - E g_i(W_{i,1})) = 0.$$

Proof. We are going to apply Corollary 1 in [44]. The δ coefficient of a stochastic matrix $P = (p(i,j))$ is defined as

$$\delta(P) = \frac{1}{2} \sup_{i,k} \sum_{j} |p(i,k) - p(k,j)|$$

In Lemma 4.1.1 we prove the strong ergodicity of the Markov chain $\{W_{n,1}\}$, so, the δ coefficient of the Corollary tends to 0 as is demanded (actually, this coefficient is 0 in all steps). Assuming $|g_n(W_{n,1})| \leq M$, all the conditions are fulfilled and the strong law follows. □

This proposition is very useful to prove some limit theorems for some moments of $\{W_{n,1}\}$.

Corollary 4.1.1. *Consider the non-homogeneous Markov Chain* $\{W_{n,1}\}$ *and assume that* $f_n \uparrow \infty$ *when* $n \to \infty$*, then*

$$\frac{\sum_{i=1}^{n} W_{i,1}}{n} \quad \to \quad w, \ a.s. \ when \ n \to \infty \tag{4.2}$$

$$\frac{\sum_{i=1}^{n} (W_{i,1})^2}{n} \quad \to \quad w^2 + \frac{w}{2}, \ a.s. \ when \ n \to \infty\,. \tag{4.3}$$

4.2 RBI for large samples

As stated in Chapter 1 and Chapter 3, when a population model is not appropriate to make inference, RBI appears as a plausible alternative. We are going to consider the test statistic presented in the previous chapter,

$$S_n = \sum_{i=1}^{n} a_i \tilde{\delta}_i, \tag{4.4}$$

a version of a rank statistic for binary responses, as in [60]. In the Chapter 3, the work was focused on the exact properties of the statistic, in this chapter, however, the study is about the asymptotic behavior of the statistic.

In order to examine asymptotic properties of S_n, a very common modification is done. By means of a change in the coefficients, S_n is converted into a martingale. This change gives a stronger stochastic structure to manage, but in return, the coefficients become more difficult.

Lemma 4.2.1. *Assume that the Klein design has been applied to a set of n patients and the responses are A_n. Then S_n can be written as follows:*

$$S_n = \sum_{j=1}^{n} b_{jn} \left(\tilde{\delta}_j - E[\tilde{\delta}_j | \mathcal{F}_{j-1}] \right), \tag{4.5}$$

where $\{b_{jn}\}_{j=1,\dots,n}$ is a sequence of coefficients which depend uniquely on A_n as stated explicitly in (4.8).

Proof. Observe that as $W_{j,1} = \omega + \sum_{i=1}^{j} \tilde{\delta}_i (a_i - 1)$, then

$$\begin{aligned} E[\tilde{\delta}_j | W_{j-1,1}] &= \frac{W_{j-1,1}}{2\omega} - \left(1 - \frac{W_{j-1,1}}{2\omega}\right) \\ &= \frac{W_{j-1,1}}{\omega} - 1 \\ &= \frac{1}{w} \sum_{i=1}^{j-1} \tilde{\delta}_i (a_i - 1). \end{aligned} \tag{4.6}$$

Hence

$$\begin{aligned} &\sum_{j=1}^{n} b_{jn} \left(\tilde{\delta}_j - E[\tilde{\delta}_j | \mathcal{F}_{j-1}] \right) = \\ &= \sum_{j=1}^{n} b_{jn} \tilde{\delta}_j + \frac{1}{w} \sum_{j=1}^{n} b_{jn} \sum_{k=1}^{j-1} \tilde{\delta}_k (1 - a_k) \\ &= \sum_{j=1}^{n} b_{jn} \tilde{\delta}_j + \frac{1}{w} \sum_{k=1}^{n-1} \left(\sum_{j=k+1}^{n} b_{jn} \right) \tilde{\delta}_k (1 - a_k) \\ &= b_{nn} \tilde{\delta}_n + \sum_{j=1}^{n-1} \left(b_{jn} + \frac{1 - a_j}{w} \sum_{k=j+1}^{n} b_{kn} \right) \tilde{\delta}_j \end{aligned} \tag{4.7}$$

Let A_n and B_n be, respectively, the column vectors $(a_1, a_2, \dots, a_n)^t$ and $(b_{1n}, b_{2n}, \dots, b_{nn})^t$. Let C_n be the upper triangular matrix of order n with $C_n(i,i) = 1$ and $C_n(i,j) = \dfrac{1 - a_i}{w}$

for $i = 1, \ldots, n$ and $j > i$. If we equate the coefficients of $\tilde{\delta}_i$ in (4.7) and S_n, we have, for each value n,

$$A_n = C_n B_n.$$

We observe that C_n^{-1} is, also, an upper triangular matrix of order n with $C_n^{-1}(i,i) = 1$ and $C_n^{-1}(i,i+1) = -\dfrac{1-a_i}{w}$ for $i = 1, \ldots, n-1$ and $C_n^{-1}(i,j) = \left(1 - \dfrac{1-a_{j-1}}{w}\right) C_n^{-1}(i,j-1)$ for $j > i+1$. So that, the coefficients $\{b_{nj}\}$, $j = 1, \ldots, n$, depend uniquely on the values of A_n because

$$B_n = C_n^{-1} A_n$$

for each value n, we have that $b_{nn} = a_n$ and for $i = 1, \ldots, n-1$,

$$b_{in} = \begin{cases} 1, & \text{if } a_i = 1 \\ \displaystyle\sum_{j=i+1}^{n} \frac{-1}{w}(1 - \frac{1}{w})^{f_{j-1} - f_i} a_j, & \text{if } a_i = 0. \end{cases} \tag{4.8}$$

□

We can use martingale techniques to study the asymptotic distribution of S_n considering

$$M_{k,n} = \sum_{j=1}^{k} b_{jn} \left(\tilde{\delta}_j - E[\tilde{\delta}_j | \mathcal{F}_{j-1}] \right), \tag{4.9}$$

where, from Lemma 4.2.1, $M_{n,n} = S_n$. The set $\{M_{k,n}, \mathcal{F}_{kn} = \mathcal{F}_k, k = 1, \ldots, n \geq 1\}$ is a square integrable triangular array of martingales with mean 0. Our target now is to obtain an appropriate normalizing sequence $\{k_n\}$ such that

$$\frac{S_n}{\sqrt{k_n}} \to \mathcal{N}(0, \sigma^2) \, . \tag{4.10}$$

This is not an direct task, because the computation of the variance of S_n in terms of the coefficients B_n can be complicated, due to the relationship between both set of coefficients. A manageable expression of the variance of S_n with the B_n coefficients would be desired. The results in Proposition 4.1.2 about the covariance structure of the $\{\tilde{\delta}_n\}$ process becomes very useful.

Proposition 4.2.1. *Assume that the Klein design has been applied to a set of n patients and the responses are A_n. Assume also that $s_n = n - f_n \uparrow \infty$ as $n \to \infty$. For $k_n = \sum_{j=1}^{n} b_{jn}^2$ we have that*

$$Var\left(\frac{S_n}{\sqrt{k_n}}\right) \to 1 - \frac{1}{2\omega} \quad \text{when } n \to \infty.$$

Proof. From Lemma 4.2.1 and Proposition 4.1.2 c) we have that

$$\begin{aligned} Var[S_n] &= B_n^t \Sigma_{\tilde{\boldsymbol{\delta}}_n - E[\tilde{\boldsymbol{\delta}}_n | W_{n-1}]} B_n \\ &= B_n^t diag(d_1, \ldots, d_n) B_n \\ &= \sum_{j=1}^{n} d_j b_{jn}^2 \end{aligned} \tag{4.11}$$

where $d_i = 1 - \dfrac{Var[W_{i-1,1}]}{w^2}$, $i = 1, \ldots, n$ because $\tilde{\boldsymbol{\delta}}_n = 2\boldsymbol{\delta}_n - 1$.

From the definition of b_{jn} in (4.8), if $a_j = 1$ then $b_{jn}^2 = 1$, so if $s_n = \sum_{j=1}^{n} a_j$ grows to infinite, $k_n = \sum_{j=1}^{n} b_{jn}^2$ also does. From Proposition 4.1.2 we know that $d_i \to 1 - \frac{1}{2w}$. Then, using Toeplitz lemma, we finally have

$$Var\left[\frac{S_n}{\sqrt{k_n}}\right] = \frac{\sum_{j=1}^{n} b_{jn}^2 d_j}{\sum_{j=1}^{n} b_{jn}^2} \longrightarrow 1 - \frac{1}{2\omega}.$$

□

Remark 4.2.1. It is worth observing that, for each n, $Var[S_n] = A_n^t \Sigma_{\tilde{\boldsymbol{\delta}}_n} A_n$, and then, $Var[S_n] = 4A_n^t \Sigma_{\boldsymbol{\delta}_n} A_n$. So, we also have an expression for the variance in terms of the responses, A_n, and the covariance matrix $\Sigma_{\boldsymbol{\delta}_n}$ which is already known from Proposition 4.1.2. Therefore, if C_n is the matrix introduced in Lemma 4.2.1 and $G_n = \sqrt{(diag(d_1, \ldots, d_n))} C_n^{-1}$, then $\Sigma_{\tilde{\boldsymbol{\delta}}_n} = G_n^t G_n$ and we have another expression for the variance,

$$Var[S_n] = A_n^t \Sigma_{\tilde{\boldsymbol{\delta}}_n} A_n = (G_n A_n)^t (G_n A_n) \, .$$

The next theorem is a central limit theorem for the test statistic S_n. Due to the modification of the statistic done in Lemma 4.2.1, we can appeal to Theorem 3.2 and

Corollary 3.2 in [30]. This theorem is a central limit theorem for martingale triangular arrays, under some conditions for the martingale difference process. These conditions require to add an extra condition about the $\{b_{jn}\}$ coefficients. The effect of this condition will be discussed later. Lets consider the notation $b_n = \Theta(a_n)$, as usual, that there exists constants k_1 and k_2 such that, $\forall n > n_0$, $k_1 a_n \leq b_n \leq k_2 a_n$.

Theorem 4.2.1. *Assume that as $n \to \infty$, $f_n \uparrow \infty$ and $s_n = n - f_n \uparrow \infty$. Assume also that*

$$\frac{b_{jn}^2}{\sum_{j=1}^{n} b_{jn}^2} = \Theta\left(\frac{1}{n}\right) \tag{4.12}$$

then

$$\frac{S_n}{\sqrt{\sum_{j=1}^{n} b_{jn}^2}} \to \mathcal{N}\left(0, \sqrt{1 - \frac{1}{2\omega}}\right). \tag{4.13}$$

Proof. For each n, we consider $[n] = max_{k=1,\ldots,n}\{k : b_{kn} \neq 0\}$. Since $s_n \to \infty$, $a_n = 1$ cases appears in infinity times, thus so does $b_{kn} = 1$ and $[n]$ diverges.

Let $\mathcal{F}_k = \sigma(W_{1,1}, \ldots, W_{k,1})$ and $M_{k,n}$ as in (4.9)

$$M_{k,n} = \sum_{j=1}^{k} b_{jn}\left(\tilde{\delta}_j - E[\tilde{\delta}_j|\mathcal{F}_{j-1}]\right).$$

Is already known that $M_{n,n} = M_{[n],n} = S_n$ and $\{M_{k,n}, \mathcal{F}_{kn} = \mathcal{F}_k, k = 1, \ldots, [n] \geq 1\}$ is a square integrable triangular array of martingales with mean 0.

For each n and $i \leq [n]$ we have that $M_{i,n} = \sum_{k=1}^{i} X_{kn}$, where

$$X_{kn} = \frac{b_{kn}(\tilde{\delta}_k - E[\tilde{\delta}_k|\mathcal{F}_{j-1}])}{\sqrt{\sum_{j=1}^{n} b_{jn}^2}}.$$

We appeal now to Theorem 3.2 and Corollary 3.2 in [30] to obtain a central limit theorem for $\{M_{k,n}\}$. We have to check the following conditions

$$\max_k |X_{k,n}| \to 0, \quad \text{in probability,} \tag{4.14}$$

$$\sum_k X_{k,n}^2 \to \eta^2, \quad \text{in probability,} \tag{4.15}$$

$$E\left(\max_k X_{k,n}^2\right) \text{ is bounded,} \tag{4.16}$$

where η^2 is a positive random variable.

The sequence $\{\tilde{\delta}_k - E[\tilde{\delta}_i|W_{k-1,1}]\}$ is bounded and $\frac{b_{jn}^2}{\sum_{j=1}^n b_{jn}^2}$ is also bounded, so (4.16) holds. Using also the bound for $\{\tilde{\delta}_k - E[\tilde{\delta}_i|\mathcal{F}_{j-1}]\}$ and the condition $\frac{b_{jn}^2}{\sum_{j=1}^n b_{jn}^2} = \Theta\left(\frac{1}{n}\right)$, (4.14) also holds.

To check (4.15), we consider, for each n, the auxiliar random variable $\tilde{X}_{k,n} = \frac{1}{\sqrt{n}}(\tilde{\delta}_k - E[\tilde{\delta}_k|\mathcal{F}_{j-1}])$, for $k = 1, \ldots, n$. Observe that $\tilde{\delta}_n = 2\delta_n - 1$ and then

$$\sum_{k=1}^n \tilde{X}_{k,n}^2 = \frac{4}{n}\sum_{k=1}^n \left(\delta_k - 2\frac{W_{k-1,1}}{2w}\delta_k + \left(\frac{W_{k-1,1}}{2w}\right)^2\right) \tag{4.17}$$

On the other hand, we have from the Borel-Cantelli lemma for Martingales and Corollary 4.1.1 that

$$\frac{\sum_{k=1}^n \delta_k}{n} \to w, \quad \frac{\sum_{k=1}^n \delta_k W_{k-1,1}}{n} \to \frac{w}{2} + w^2 \text{and} \quad \frac{\sum_{k=1}^n (W_{k-1,1})^2}{n} \to \frac{w}{2} + w^2 \quad a.s.$$

Therefore, from (4.17), we conclude that $\sum_{k=1}^n \tilde{X}_{k,n}^2$ is convergent.

Observe that from (4.2) and from the definition of $[n]$, for each $1 \leq k \leq [n]$, there exists $n_0 \geq 1$ and positive constants K_1, $K_2 > 0$ such that for each $n \geq n_0$ we have that $K_1/n \leq \frac{b_{kn}^2}{\sum_{j=1}^n b_{jn}^2} \leq K_2/n$. From Theorem 12.4 in [42] it follows that the series $\sum_{k=1}^n \tilde{X}_{k,n}^2$ and $\sum_{k=1}^n X_{k,n}^2 = \sum_{k=1}^{[n]} X_{k,n}^2$ converge or diverge together. So that, condition (4.16) holds and, then, S_n converges to a normal distribution with zero mean and variance η^2 which satisfies

$$\eta^2 = \lim_{n\to\infty} Var\left[\frac{S_n}{\sqrt{\sum_{j=1}^n b_{jn}^2}}\right].$$

Now, from Proposition 4.2.1, $\eta^2 = 1 - \frac{1}{2w}$ and the theorem is proved. □

It is not difficult to find out examples where condition (4.2) does not hold. For instance, assume that the sequence of responses is always successful except for the first patient, that is, $a_1 = 0$ and $a_i = 1$ for $i \geq 2$. Then, for each n, $b_{n1} = -(n-1)/w$ and $b_{ni} = 1$ for $i \geq 2$. So that, condition (4.2) does not hold. Condition (4.2) appears too stringent, but the following corollary shows a situation where it holds.

Corollary 4.2.1. *Let $\{a_i\}_{i\geq 1}$ be a sequence of zeros and ones. Assume that the length of any run of ones is bounded by a value M. Then, for any $j = 1, \ldots, n$,*

$$\frac{b_{jn}^2}{\sum_{j=1}^{n} b_{jn}^2} = \Theta\left(\frac{1}{n}\right)$$

where $\{b_{jn}\}_{j=1,\ldots,n}$ are the coefficients of (4.8).

Proof. For each value n, fix a value $i = 1 \ldots n-1$ and denote R_k the length of the kth run of ones after the position i of A_n and L_k the length of the kth run of zeros before the kth run of ones after the position i of A_n. Let t the number of runs of ones from the ith position until the nth position. Then, if $a_i = 0$, from (4.8) we have

$$\begin{aligned}
|b_{in}| &= \frac{1}{w}\left(R_1 + (1-\frac{1}{w})^{L_1} R_2 + \cdots + (1-\frac{1}{w})^{\sum_{i=1}^{t-1} L_i} R_t\right) \\
&\leq \frac{M}{w}\left(1 + (1-\frac{1}{w})^{L_1} + \cdots + (1-\frac{1}{w})^{\sum_{i=1}^{t-1} L_i}\right) \\
&\leq M\left(1 - (1-\frac{1}{w})^{n+1}\right) \\
&\leq M.
\end{aligned}$$

So for each n and for any $i = 1, \ldots, n$, $|b_{in}|$ is a bounded value and the result follows. □

Observe that condition (4.2) deals with a non finite sequence of responses. However, in a clinical trial, a finite number of patients participate and, so, we have a finite sequence of responses A_n. Although Corollary 4.2.1 gives a quite realistic condition, it also relies on the assumption that an infinite sequence of responses is available. In [60] the same situation happens and to determine how a typical response sequence behaves, A_n is considered a realization of a Bernoulli trial of length n, that is, a sequence of random variables Y_i, $i = 1, \ldots, n$ independent and identically distributed with $P(Y_i = 1) = p$ and $P(Y_i = 0) = q$.

For each n, we define B_{nj} as b_{nj} in (4.8) via the sequence $\{Y_i\}$, $i = 1, \ldots, n$ instead

of A_n. Then, for each n and $1 \leq j \leq n$

$$B_{nj} = Y_j - \frac{1-Y_j}{w} \sum_{i=j+1}^{n} \left(1 - \frac{1}{w}\right)^{\sum_{r=j+1}^{i-1}(1-Y_r)} Y_i \tag{4.18}$$

where $\sum_a^b = 0$ when $b < a$.

It is not difficult to see from (4.18) that for each n and $1 \leq j \leq n$

$$B_{nj} = Y_j - \frac{1-Y_j}{w} \sum_{r=1}^{F_j} L_r^j (1 - \frac{1}{w})^{r-1}.$$

where, F_j is the number of failures from the $(j+1)$-th patient up to, and including, the n-th patient, and L_r^j is the number of successes between the r-th and $(r+1)$-th failures, so $L_r^j = 0$ when both zeros are consecutive. If $Y_n = 1$, then $L_{F_j}^j$ is equal to the number of ones between the last failure and n and otherwise, is equal to zero.

Let L_n be the longest run of successes in n independent Bernoulli trials with success probability p. It is well-known that $L_n / log_{1/p} n \to 1$, almost surely, see, for instance, section 2.2.5 in [4]. On the other hand, $\sum_{j=1}^n B_{nj}^2 \geq \sum_{j=1}^n Y_j$ and, for a constant C $B_{nj}^2 < C L_n^2$. So that, for any $0 < p < 1$,

$$\frac{max_{1 \leq j \leq n} B_{nj}^2}{\sum_{j=1}^{n} B_{nj}^2} \to 0, \quad a.s.$$

We only have to elucidate if, for any $0 \leq p \leq 1$, condition (4.2) holds, that is, if the sequence $\{B_n\}$ is $\Theta\left(\frac{1}{n}\right)$, where, for each n,

$$B_n = \frac{max_{1 \leq j \leq n} B_{nj}^2}{\sum_{j=1}^{n} B_{nj}^2}.$$

For $w = 1$, it is immediate to prove that $\{B_n\}$ is $\Theta\left(\frac{1}{n}\right)$ for any $0 < p < 1$. However, small values of w entail large asymptotic selection bias as measured in [26]. When $w \geq 2$, intensive simulation studies have been carried out and support that $\{B_n\}$ is $\Theta\left(\frac{1}{n}\right)$.

Lemma 4.2.2. *The triangular array of random variables* $\{B_{nj} : n \geq 1,\ 1 \leq j \leq n\}$ *satisfies the following properties:*

a) For each n *and* $1 \leq j \leq n$ *we have that*

$$E[B_{nj}] = p\left(1 - \frac{q}{w}\right)^{n-j}.$$

b) Let $A_1 = p + q\left(1 - \frac{1}{w}\right)^2$ *and* $A_2 = 1 - \frac{q}{w}$. *For each* n *and* $1 \leq j \leq n$ *we have that*

$$E[B_{nj}^2] = p + \frac{p}{w^2}\left(\frac{1 - A_1^{n-j}}{1 - A_2^2} + 2pw\sum_{r=j+1}^{n-1} A_1^{r-j-1}(1 - A_2^{n-r})\right).$$

c) When $n \to \infty$,

$$\sum_{j=1}^{n} E[B_{nj}] \to \frac{pw}{q}.$$

d) When $n \to \infty$,

$$\sum_{j=1}^{n} E[B_{nj}]/n \to 0 \quad \textit{and} \quad \sum_{j=1}^{n} E[B_{nj}^2]/n \to p + \frac{pq + 2p^2w}{w^2(1 - A_1)}$$

Proof. Taking expectations in the definition of B_{nj} in (4.18), we get

$$\begin{aligned} E[B_{nj}] &= p - \frac{qp}{w}\sum_{i=j+1}^{n} E\left[(1 - \frac{1}{w})^{\sum_{r=j+1}^{i-1}(1-Y_r)}\right] \\ &= p - \frac{qp}{w}\sum_{i=j+1}^{n}\prod_{r=j+1}^{i-1} E\left[\left(1 - \frac{1}{w}\right)^{(1-Y_r)}\right]. \end{aligned}$$

Considering that Y_r are Bernoulli variables, $E\left[\left(1 - \frac{1}{w}\right)^{(1-Y_r)}\right] = p + q\left(1 - \frac{1}{w}\right) = 1 - \frac{q}{w}$, so

$$\begin{aligned} E[B_{nj}] &= p - \frac{qp}{w}\sum_{i=j+1}^{n}\left(1 - \frac{q}{w}\right)^{i-j-1} = p - \frac{qp}{w}\sum_{i=0}^{n-j-1}\left(1 - \frac{q}{w}\right)^{i} \\ &= p - \frac{qp}{w}\,\frac{1 - \left(1 - \frac{q}{w}\right)^{n-j}}{1 - \left(1 - \frac{q}{w}\right)} = p\left(1 - \frac{q}{w}\right)^{n-j}, \end{aligned}$$

and part a) is proven. For part b), using (4.18) we can express B_{nj}^2 as

$$B_{nj}^2 = Y_j - \frac{1 - Y_j}{w^2}\left(\sum_{r=j+1}^{n} l_r\right)^2,$$

where $l_r = Y_r \left(1-\frac{1}{w}\right)^{\sum_{s=j+1}^{r-1}(1-Y_s)} = Y_r \prod_{s=j+1}^{r-1} \left(1-\frac{1}{w}\right)^{1-Y_s}$.

Now,

$$E[B_{nj}^2] = p + \frac{q}{w^2}\left[\sum_{r=j+1}^{n} E\left[l_r^2\right] + 2 \sum_{t>nr=j+1,n} E[l_r l_t]\right], \tag{4.19}$$

$$\begin{aligned} E\left[l_r^2\right] &= E\left[Y_r \prod_{s=j+1}^{r-1} \left(1-\frac{1}{w}\right)^{2(1-Y_s)}\right] \\ &= p \prod_{s=j+1}^{r-1} \left[p + q\left(1-\frac{1}{w}\right)^2\right] = p\left[p + q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1}, \end{aligned}$$

taking into account the independence between Y_r and the product. Computing all the summation,

$$\begin{aligned} \sum_{r=j+1}^{n} E\left[l_r^2\right] &= p \sum_{r=j+1}^{n} \left[p + q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1} = p \sum_{r=0}^{n-j-1} \left[p + q\left(1-\frac{1}{w}\right)^2\right]^{r} \\ &= p \frac{1-\left[p+q\left(1-\frac{1}{w}\right)^2\right]^{n-j}}{1-\left[p+q\left(1-\frac{1}{w}\right)^2\right]} = \frac{p}{q} \frac{1-\left[p+q\left(1-\frac{1}{w}\right)^2\right]^{n-j}}{1-\left(1-\frac{1}{w}\right)^2} \end{aligned} \tag{4.20}$$

We follow a similar procedure for $E[l_r l_t]$. Assuming $t > r$,

$$\begin{aligned} E[l_r l_t] &= E\left[Y_r \prod_{s=j+1}^{r-1} \left(1-\frac{1}{w}\right)^{(1-Y_s)} Y_t \prod_{s=j+1}^{t-1} \left(1-\frac{1}{w}\right)^{(1-Y_s)}\right] \\ &= E\left[Y_r \left(1-\frac{1}{w}\right)^{r} \prod_{s=j+1}^{r-1} \left(1-\frac{1}{w}\right)^{2(1-Y_s)} Y_t \prod_{s=r+1}^{t-1} \left(1-\frac{1}{w}\right)^{(1-Y_s)}\right] \\ &= E\left[Y_r \left(1-\frac{1}{w}\right)^{Y_r}\right] E\left[\prod_{s=j+1}^{r-1} \left(1-\frac{1}{w}\right)^{2(1-Y_s)}\right] E\left[Y_t \prod_{s=r+1}^{t-1} \left(1-\frac{1}{w}\right)^{(1-Y_s)}\right], \end{aligned}$$

due to the independence of the different responses. Calculating the expectations as above,

$$\begin{aligned} E[l_r l_t] &= p\left[p + q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1} p\left[p + q\left(1-\frac{1}{w}\right)\right]^{t-r-1} \\ &= p^2\left[p + q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1} \left(1-\frac{q}{w}\right)^{t-r-1}. \end{aligned}$$

Calculating the whole summation,

$$\begin{aligned}
\sum_{r=j+1}^{n-1}\sum_{t=r+1}^{n} E[l_r l_t] &= \sum_{r=j+1}^{n}\sum_{t=r+1}^{n} p^2\left[p+q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1}\left(1-\frac{q}{w}\right)^{t-r-1} \\
&= \sum_{r=j+1}^{n-1} p^2\left[p+q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1}\sum_{t=0}^{n-r-1}\left(1-\frac{q}{w}\right)^{t} \\
&= \sum_{r=j+1}^{n-1} p^2\left[p+q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1}\frac{1-\left(1-\frac{q}{w}\right)^{n-r}}{\frac{q}{w}} \\
&= \frac{wp^2}{q}\sum_{r=j+1}^{n-1}\left[p+q\left(1-\frac{1}{w}\right)^2\right]^{r-j-1}\left[1-\left(1-\frac{q}{w}\right)^{n-r}\right] \quad (4.21)
\end{aligned}$$

Plugging (4.21) and (4.20) into (4.19) and denoting $A_1 = p + q\left(1-\frac{1}{w}\right)^2$ $A_2 = 1-\frac{q}{w}$ we get the expression in b),

$$E[B_{nj}^2] = p + \frac{p}{w^2}\left[\frac{1-A_1^{n-j}}{1-\left(1-\frac{1}{w}\right)^2} + 2wp\sum_{r=j+1}^{n-1} A_1^{r-j-1}\left[1-A_2^{n-r}\right]\right]$$

For part c),

$$\sum_{j=1}^{n} E[B_{nj}] = \sum_{j=1}^{n} p\left(1-\frac{q}{w}\right)^{n-j} = p\sum_{j=1}^{n-1}\left(1-\frac{q}{w}\right)^{j}$$

and taking limits the result is direct. The last part follows straightforwardly noting that $E[B_{nj}] \to 0$ and taking a cesaro-type limit. □

By assuming that a Bernoulli trial gives the set of responses, we can give more explicit expressions about the random variable S_n.

Proposition 4.2.2. *Consider that a Bernoulli trial of length n and success probability p generates the sequence of responses in a clinical trial where the allocations are made with a Klein design with, initially, $w > 1$ balls of each color. Then, as $n \to \infty$*

$$\frac{E\left[Var[S_n]\right]}{n} \to \left(p + \frac{pq+2p^2w}{w^2(1-A_1)}\right)\left(1-\frac{1}{2w}\right)$$

Proof. Let $\mathcal{Y}_n = \sigma(Y_1,\ldots,Y_n)$. Observe that A_n is a specific realization of the Bernoulli trial. Let $D_j = \frac{w}{2}(1-\frac{2}{w})^{\sum_{i=1}^{j-1}(1-Y_i)}$. From (4.11) we have that

$$E[Var[S_n]] = \sum_{j=1}^{n} E[D_j]E[B_{nj}^2] \quad (4.22)$$

where the equality follows from the fact that D_j is $\mathcal{Y}_{j-1}$-measurable and B_{nj}^2 is $\sigma(Y_j, ..., Y_n)$-measurable.

It is easy to see that $E[D_j] = 1 - \frac{1}{2w}\left(1 - \left(1 - \frac{2}{w}\right)^j\right)$ and from Lemma 3.2 the result follows straightforwardly. □

Remark 4.2.2. The study of central limit theorems for non-homogeneous chains is not a closed problem and, up to our knowledge, there is not in the literature any theorem in which our chain fits. On the topic of central limit theorems for non-homogeneous Markov chains, [55] is an updated reference that explains the state of the art and gives new sufficient conditions for the existence of a central limit theorem for

$$\sum_{i=1}^{n} W_{i,1} / Var\left[\sum_{i=1}^{n} W_{i,1}\right].$$

Those conditions are based on the maximal coefficient of correlation $\rho_{n,1}$. For $\{W_{n,1}\}$, following [55] we have that

$$\rho_{n,1} = max_{1 \leq k \leq n-1} \rho(W_{k,1}, W_{k+1,1}).$$

Assuming the conditions of Proposition 4.1.2 we have to compute $Cov(W_{k,1}, W_{k+1,1})$. Using the decomposition of $W_{k,1}$ in (4.1),

$$Cov(W_{k,1}, W_{k+1,1}) = Var(W_{k,1}) - Cov\left(W_{k,1}, (1 - a_{k+1})\,\tilde{\delta}_{k+1}\right)$$

Using Proposition 2.2.1

$$\begin{aligned} Cov\left(W_{k,1}, (1 - a_{k+1})\,\tilde{\delta}_{k+1}\right) &= Cov\left(W_{k,1}, E\left[(1 - a_{k+1})\,\tilde{\delta}_{k+1} | \mathcal{F}_k\right]\right) \\ &= Cov\left(W_{k,1}, (1 - a_{k+1})\,\frac{W_{k,1} - 1}{w}\right) \\ &= \frac{1 - a_{k+1}}{w} Var(W_{k,1}) \end{aligned}$$

So, we get the next relationship,

$$Cov(W_{k,1}, W_{k+1,1}) = Var(W_{k,1})\left(1 - \frac{1 - a_{k+1}}{w}\right).$$

As if $a_k = 1$ then $Var(W_{k,1}) = Var(W_{k-1,1})$, from the behavior of the chain, in case of success the distribution does not change. Then, assume that for some value n_0, $a_{n_0+1} = 1$. This condition is not stringent, because in some step a success is expected. Therefore, we obtain that $\rho(W_{n_0,1}, W_{n_0+1,1}) = 1$. So that, $\rho_{n,1} = 1$ for any $n \geq n_0$; and the conditions on [55] to obtain a central limit theorem for $\{W_{n,1}\}$ do not hold. So, the most recent central limit theorem for non-homogeneous Markov chains is not valid for our purposes.

Remark 4.2.3. Remember that

$$S_n = W_{n,1} - w + \Delta_n$$

where $\Delta_n = 2N_{n,1} - n$ and $\{W_{n,1} - w\}$ is a bounded sequence, given an appropriate sequence $\{k_n\}$ of positive numbers which grows to infinity we obtain

$$\frac{S_n}{\sqrt{k_n}} = \frac{W_{n,1} - w}{\sqrt{k_n}} + \frac{\Delta_n}{\sqrt{k_n}},$$

so that, in the conditions of Theorem 4.2.1, we obtain a central limit theorem for

$$\frac{2n}{\sqrt{k_n}} \left(\frac{N_{n,1}}{n} - \frac{1}{2} \right).$$

Now, the non-homogeneous Markov Chain $\{W_{n,1}\}$ has a Central Limit Theorem, following the same reasoning of Proposition 4 in [26] where a central limit theorem is obtained for the non-conditioned process $\{W_{n,1}\}$.

4.3 Simulation study

The quality of the normal approximation of the p-values has been studied via Monte Carlo simulation and Table 4.3 gives a summary of the results. For a sequence of patients' responses of length n, generated with a Bernoulli distribution of parameter p, the design is run 15000 times, generating 15000 values of $\frac{S_n}{\sqrt{k_n(1-\frac{1}{2\omega})}}$. The percentage, $r(i)$, of these values smaller than the i-percentile of the normal standard distribution is calculated. To obtain the precision of the estimator, this procedure is repeated 1000 times for $i =$

80, 90, 95 and 99 and for sample sizes $n = 50$, 150 and 500. The mean values (and standard deviations) of $r(i)$ are displayed in Table 4.3.

These values give us an accurate information about how close are the tails of both distributions. We observe that the proportion $r(i)$ is close to i, even for $n = 50$ and for the most extreme percentile $i = 99$. As could be expected, the approximation is better for greater values of n.

Table 4.1: Proportion of values of the test-statistic $S_n/\sqrt{k_n(1-\frac{1}{2\omega})}$ lower than the i-percentile of the standard normal distribution.

	$i = 80$			$i = 90$		
p	$n = 50$	$n = 150$	$n = 500$	$n = 50$	$n = 150$	$n = 500$
0.9	0.7914	0.7939	0.7971	0.8934	0.8956	0.8981
	(0.0201)	(0.0084)	(0.0044)	(0.0135)	(0.0056)	(0.0031)
0.5	0.7945	0.7978	0.7996	0.8957	0.8983	0.8996
	(0.0279)	(0.0142)	(0.0080)	(0.0180)	(0.0091)	(0.0052)
0.1	0.7917	0.7985	0.7976	0.8930	0.9004	0.8989
	(0.0522)	(0.0413)	(0.0215)	(0.0367)	(0.0260)	(0.0146)
	$i = 95$			$i = 99$		
p	$n = 50$	$n = 150$	$n = 500$	$n = 50$	$n = 150$	$n = 500$
0.9	0.9466	0.9476	0.9493	0.9899	0.9902	0.9904
	(0.0081)	(0.0036)	(0.0021)	(0.0023)	(0.0011)	(0.0009)
0.5	0.9482	0.9493	0.9500	0.9908	0.9904	0.9901
	(0.0106)	(0.0055)	(0.0032)	(0.0027)	(0.0016)	(0.0011)
0.1	0.9521	0.9500	0.9497	0.9917	0.9909	0.9905
	(0.0205)	(0.0161)	(0.0083)	(0.0059)	(0.0038)	(0.0023)

Chapter 5

Covariates and the Klein urn design

Covariates or prognostic factors can give important information about patients and can have influence in patient's response to the treatment. An imbalance between covariate levels could produce a bias in the study. Even if randomization itself protects against these imbalances, see [63] Chapter 5, due to the asymptotic nature of this property, in moderate sample sizes some imbalance could appear, biasing the study. Different solutions have been presented in order to avoid or mitigate this problem. For instance, a technique deals with the inclusion of covariates in the design of the trial. In Chapter 4 of [63] and in [65] good reviews of the use of covariates in adaptive designs can be seen.

If we want to use the patient's covariates in the allocation rule in an urn-based design it seems natural to make a stratification and assign a different urn to each stratum. The Klein urn design has been introduced as a clinical trial design without taking into account covariates. In this chapter we study how to introduce covariate information with the goal of improving the inferential procedure. A stratified version of the Klein urn design is presented here, and some basic properties are studied. Besides, we consider a model that relates the observed responses with the treatment applied and with some patient's

covariates. The main difficulty is the dependence between the error terms in the model due to the adaptive allocation process. So, consistency and asymptotic normality of the estimators need to be assessed. Finally, a study is made via simulation, comparing a classical Wald test, a randomization test and a test based on the model.

5.1 The stratified response-adaptive Klein urn design

A clinical trial with two treatments is considered. We assume that patients arrive sequentially and their covariate information is available before they are allocated to a treatment. With this information the stratification is done. Each stratum can be seen as a combination of levels of relevant covariates. For instance, it could be assumed that there are M covariates (or prognostic factors) $\mathbf{H} = (H_1, \ldots, H_M)$, with $m_1, \ldots, m_M$ different levels, respectively, and therefore the number of strata is $K = \prod_{i=1}^{M} m_i$. The indicator function of the n-th patient's stratum is denoted by

$$\pi_{nk} = \begin{cases} 1, & n\text{-th patient belongs to the } k\text{-th strata;} \\ 0, & \text{otherwise.} \end{cases}$$

We assume that $\{\pi_{nk}\}$ are independent Bernoulli random variables with mean π_k, where $\pi_k > 0$, $k = 1, \ldots, K$ and $\sum_{k=1}^{K} \pi_{nk} = 1$. Patients are allocated to a treatment using a different urn for each stratum. When a patient that belongs to stratum k arrives, a ball is drawn from the urn k, and the patient is allocated to the corresponding treatment. For the sake of clarity, in this chapter the indicator function of the n-th patient's treatment is δ_{nj}, $j = 1, 2$. Relating with the indicators used in the previous chapters, $\delta_{n1} = \delta_n$ and $\delta_{n2} = 1 - \delta_n$. Note that its distribution depends on the stratum of the patient.

We assume that the response is observed before the next patient arrives. We denote by Z_{nkj} the response of the n-th patient to the treatment j, when this patient belongs to stratum k. Note that only one of the possible $2 \times K$ responses is really observed. If this response is dichotomous (success or failure) with success probability p_{kj}, we denote the failure indicator as $F_{nkj} = 1 - Z_{nkj}$. In general, the indicator of failure can be defined

as $F_{nkj} = g(X_{nkj})$ were the function g can be seen as a discretization to failure/success of, maybe, a continuous response X_{nkj}. We assume that $\{F_{nkj}\}$ are Bernoulli random variables with mean $q_{kj} = 1 - p_{kj}$. The replacement policy is the usual one in the Klein urn: If the treatment is a failure, we put a ball of the other type instead of the extracted ball; otherwise, we replace the extracted ball and the urn remains unchanged.

Therefore, the urn is updated with the following replacement matrix

$$\begin{array}{lll} & & \begin{array}{cc} \textit{type } 1 & \textit{type } 2 \\ \downarrow & \downarrow \end{array} \\ \text{Get type 1 ball} & \rightarrow & \\ \text{Get type 2 ball} & \rightarrow & R_i = \begin{pmatrix} -F_{nk1} & F_{nk1} \\ F_{nk2} & -F_{nk2} \end{pmatrix} \end{array}$$

This procedure will be called stratified Klein urn design. If we have allocated n patients, we denote the number of patients in stratum k assigned to treatment j as

$$N_{nkj} = \sum_{i=1}^{n} \delta_{ij} \pi_{ik} \, .$$

We denote the total number of patients in stratum k as $N_{nk\cdot}$ and the total number of patients assigned to treatment j as $N_{n\cdot j}$,

$$N_{nk\cdot} = N_{nk1} + N_{nk2} \, ; \quad N_{n\cdot j} = \sum_{k=1}^{K} N_{nkj} \, .$$

For instance, consider a case with two covariates with two levels each, this would lead to a scenario with $2 \times 2 = 4$ strata. Here we show a displayed table of n allocations with 4 strata.

Stratum	Treatment 1	Treatment 2	Total
1	N_{n11}	N_{n12}	$N_{n1\cdot}$
2	N_{n21}	N_{n22}	$N_{n2\cdot}$
3	N_{n31}	N_{n32}	$N_{n3\cdot}$
4	N_{n41}	N_{n42}	$N_{n4\cdot}$
Total	$N_{n\cdot 1}$	$N_{n\cdot 2}$	n

Due to the heterogeneity between patients in the different strata, the probability of success also can be different. So, the behavior of the urns linked to each stratum could be different. The limit proportion of treatment 1 allocations can be computed for each stratum and using these quantities the overall treatment allocation proportion is also computed.

Proposition 5.1.1. *Assume that the stratified Klein urn design is applied. Then, for each stratum k in $0, ..., K$,*

$$\frac{N_{nk\cdot}}{n} \to \pi_k \qquad a.s.$$

$$\frac{N_{nk1}}{N_{nk\cdot}} \to a_{k1} := \frac{q_{k2}}{q_{k1} + q_{k2}} \qquad a.s.$$

$$\frac{N_{n\cdot 1}}{n} \to \sum_{k=1}^{K} a_{k1}\pi_k \qquad a.s.$$

Proof. Observe that $\{\pi_{nk}\}$ are independent Bernoulli random variables with mean π_k, for each stratum k. So, from the classical strong law of large numbers,

$$\frac{N_{nk\cdot}}{n} = \frac{\sum_{i=1}^{n} \pi_{ik}}{n} \to \pi_k \qquad a.s. \tag{5.1}$$

The probability of assigning each stratum, π_k, is strictly positive so, $N_{nk\cdot}$, the number of times that each stratum is assigned (i.e. the number of patients in each urn) goes to infinite. Moreover, as these random variables, $\{\pi_{ik}\}$, are independent, then each urn works independently from the others, and we can apply the theory of Chapter 2 in each stratum. Being q_{k1} the probability of failure of treatment 1 in the k-th stratum, Proposition 2.4 in Chapter 2 gives

$$\frac{N_{nk1}}{N_{nk\cdot}} \to \frac{q_{k2}}{q_{k1} + q_{k2}} = a_{k1}, \qquad a.s. \tag{5.2}$$

Finally, combining (5.1) and (5.2), we get the convergence of the total proportion of treatment 1 allocations,

$$\frac{N_{n\cdot 1}}{n} = \frac{\sum_{k=1}^{K} N_{nk1}}{n} = \sum_{k=1}^{K} \frac{N_{nk1}}{N_{nk\cdot}} \frac{N_{nk\cdot}}{n} \to \sum_{k=1}^{K} \frac{q_{k2}}{q_{k1} + q_{k2}} \pi_k \quad .$$

□

The stratification also permits doing stratified inference analysis, defining the usual estimators for each stratum. We denote the proportion of successes of treatment j in stratum k until patient n as

$$\hat{p}_{nkj} = \frac{1}{N_{nkj}} \sum_{i=1}^{n} Z_{ikj}\delta_{ij}\pi_{ik} \,.$$

The average of successes for treatment j can be expressed as

$$\hat{p}_{n.j} = \frac{\sum_{k=1}^{K} N_{nkj}\hat{p}_{nkj}}{N_{n.j}} \,.$$

If we consider the case with non dichotomous responses, then X_{ikj} are quantitative random variables with mean μ_{kj}. We denote the average of treatment j in stratum k until patient n as

$$\hat{\mu}_{nkj} = \frac{1}{N_{nkj}} \sum_{i=1}^{n} X_{ikj}\delta_{ij}\pi_{ik} \,.$$

Then, the average of treatment j can be expressed as

$$\hat{\mu}_{n.j} = \frac{\sum_{k=0}^{K} N_{nkj}\hat{\mu}_{nkj}}{N_{n.j}} \,.$$

In the following proposition the limit behavior of these estimators is obtained.

Proposition 5.1.2. *Assume that the stratified Klein urn design is applied. If the responses Z_{ikj} are independent Bernoulli random variables with mean p_{kj} for each stratum k in $1, ..., K$ and treatment $j = 1, 2$, then*

$$\begin{aligned}
\hat{p}_{nkj} &\to p_{kj} \qquad a.s. \\
\hat{p}_{n.j} &\to p_{.j} := \frac{\sum_{k=1}^{K} \pi_k a_{kj} p_{kj}}{\sum_{m=1}^{K} \pi_m a_{mj}} \qquad a.s. \\
\sqrt{N_{nkj}}(\hat{p}_{nkj} - p_{kj}) &\to N(0, p_{kj}q_{kj}) \quad [D] \\
\sqrt{n}(\hat{p}_{nkj} - p_{kj}) &\to N(0, p_{kj}q_{kj}/(a_{kj}\pi_k)) \quad [D]
\end{aligned}$$

If the responses X_{ikj} are independent non-binary random variables with mean μ_{kj} and

variance $Var(X_{nkj}) = \sigma^2_{kj}$ *for each stratum* k *in* $1, ..., K$ *and treatment* $j = 1, 2$*, then*

$$\begin{aligned} \hat{\mu}_{nkj} &\to \mu_{kj} \qquad a.s. \\ \hat{\mu}_{n.j} &\to \mu_{.j} := \frac{\sum_{k=1}^{K} \pi_k a_{kj} \mu_{kj}}{\sum_{m=1}^{K} \pi_m a_{mj}} \qquad a.s. \\ \sqrt{N_{nkj}}(\hat{\mu}_{nkj} - \mu_{kj}) &\to N(0, \sigma^2_{kj}) \quad [D] \\ \sqrt{n}(\hat{\mu}_{nkj} - \mu_{kj}) &\to N(0, \sigma^2_{kj}/(a_{kj}\pi_k)) \quad [D] \end{aligned}$$

Proof. As the probability of assigning stratum k, π_k $k = 1, ..., K$, is strictly positive, then $N_{nk.} \to \infty$. Applying Proposition 3 in [26] or Proposition 2.3 in Chapter 2, we know that $N_{nkj} \to \infty$ for any pair (k, j). So, in each stratum treatment 1 and 2 are applied infinite times. Theorem 1 in [11] could be applied for each stratum. Then, strong laws can be applied to these estimators and therefore,

$$\hat{p}_{nkj} \to p_{kj} \qquad a.s..$$

For the average of successes in treatment j, we have to compute the proportion of treatment 1 allocations that are in the k-th stratum and

$$\frac{N_{nkj}}{N_{n.j}} = \frac{N_{nkj}}{N_{nk.}} \frac{N_{nk.}}{n} \frac{n}{N_{n.j}} \to a_{kj}\pi_k \frac{1}{\sum_{m=1}^{K} \pi_m a_{mj}}$$

using the three convergences of Proposition 5.1.1. Then, the overall limiting proportion can be computed

$$\hat{p}_{n.j} = \frac{\sum_{k=1}^{K} N_{nkj}\hat{p}_{nkj}}{N_{n.j}} = \sum_{k=1}^{K} \frac{N_{nkj}}{N_{n.j}} \hat{p}_{nkj} \to \frac{\sum_{k=1}^{K} \pi_k a_{kj} p_{kj}}{\sum_{m=1}^{K} \pi_m a_{mj}}$$

The adaptive central limit theorem also holds for each stratum,

$$\sqrt{N_{nkj}}\,(\hat{p}_{nkj} - p_{kj}) \to N(0, p_{kj}q_{kj}) \quad [D]\,.$$

Noting that $\frac{n}{N_{nkj}} = \frac{n}{N_{nk.}} \frac{N_{nk.}}{N_{nkj}} \to \frac{1}{a_{kj}\pi_j}$ $a.s.$, if we replace $\sqrt{N_{nkj}}$ by $\sqrt{n}$, the central limit still holds adjusting the variance,

$$\sqrt{N_{nkj}}\,(\hat{p}_{nkj} - p_{kj}) \to N(0, p_{kj}q_{kj}) \quad [D]\,.$$

The results for the case of non-dichotomous responses follow with an analogous argumentation. □

Assuming a population model, conventional inference could be done. This inference could be stratum by stratum or a global test encompassing all of them. Using a test statistic that includes the covariates that have been used in the randomization procedure is recommended in [69] and [1].

Remark 5.1.1. Another way to do inference is the use of stratified randomization tests, as in Chapter 8 in [63]. The main idea is to use the randomization test presented in Chapter 3 in each stratum and then combine all these statistics in a common statistic. That is, let $S_{nk} = \sum_{i=1}^{N_{nk\cdot}} a_{i(k)}\tilde{\delta}_{ik}$ be the test statistic in the k-th stratum, being $a_{i(k)}$ the fixed response of the i-th patient in the k-th stratum. Then, a global test is defined as

$$S_n = \sum_{j=1}^{K} \omega_j S_{nj} \,,$$

where the ω_j are weights which define the importance of the different strata, where the weights ω_j are usually the proportion of patients in each stratum.

5.2 Generalized linear adaptive models

When we assume a model that relates the response with some covariates, the effect of the treatments can be included in this model. It is interesting to study how the use of an adaptive design affects the estimation of the parameters. The generalized linear models (GLM) are a general framework for this study. In such models, it is assumed that the distribution of the responses belongs to the exponential family, and these responses are related, via a link function, with a linear function of the covariates, that is, with a linear predictor. GLM were introduced in [54] and a detailed study of their properties can be seen in [48].

Dichotomous responses, success or failure, are being considered in our study, so, the canonical link function is the logistic function. However, as we are using an adaptive design to allocate patients, the classical assumption of independence or incorrelation do not hold, so, we have that the usual properties of the estimators are not guaranteed, therefore, a

more detailed study is needed.

Let Y_n be the dichotomous random variable indicating the response of the n-th patient, with $Y_n = 1$ when the treatment is successful and $Y_n = 0$ when the treatment fails. Let the natural filtration be $\mathcal{F}_{n-1} = \sigma(\mathcal{Z}_{n-1}, \mathcal{G}_n, \mathcal{H}_n)$, where $\mathcal{Z}_n$, $\mathcal{G}_n$ and $\mathcal{H}_n$ are defined as in section 1.1 . Then, Y_n is $\mathcal{F}_n$-measurable and the regressors are $\mathcal{F}_{n-1}$-measurable. We denote

$$E[Y_n|\mathcal{F}_{n-1}] = \mu(\boldsymbol{\beta}^t \boldsymbol{x}_n) = \frac{e^{\beta^t \boldsymbol{x}_n}}{e^{\beta^t \boldsymbol{x}_n} + 1} = \mu_n$$

$$Var[Y_n|\mathcal{F}_{n-1}] = v(\mu_n) = \frac{e^{\beta^t \boldsymbol{x}_n}}{(e^{\beta^t \boldsymbol{x}_n} + 1)^2} = v_n \, .$$

where $\mu(\cdot)$ is the link function, $\boldsymbol{\beta}$ is the vector of parameters and $\boldsymbol{x}_n$ is the vector of regressor variables and, therefore, $\eta = \boldsymbol{\beta}^t \boldsymbol{x}_n$ is the linear predictor. The regression relation has now the form

$$Y_n = \mu(\boldsymbol{\beta}^t \boldsymbol{x}_n) + \epsilon_n \, , \tag{5.3}$$

where $\epsilon_n = Y_n - E[Y_n|\mathcal{F}_{n-1}]$ form a martingale difference sequence.

For the linear predictor $\eta = \boldsymbol{\beta}^t \boldsymbol{x}_n$, three different models are considered: (A) assumes a linear growing effect on the levels of the strata, (B) uses a parameter for each stratum and (C) uses a parameter for each treatment-stratum interaction.

$$\eta = \begin{cases} \sum_{j=1}^{2} \mu_j \delta_{nj} + \gamma \sum_{k=1}^{K} k \pi_{nk} & (A) \, , \\ \sum_{j=1}^{2} \mu_j \delta_{nj} + \sum_{k=1}^{K} \gamma_k \pi_{nk} & (B) \, , \\ \sum_{j=1}^{2} \mu_j \delta_{nj} + \sum_{j=1}^{2} \sum_{k=1}^{K} \gamma_{kj} \delta_{nj} \pi_{nk} & (C) \, . \end{cases}$$

In order to avoid multicollinearity and to ensure the invertibility of the design matrix $D_n = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^t$, some changes have been made in models (B) and (C). In model (B), the indicators of treatment 2 and covariate K have been dropped and so the intercept α reflects this baseline status: to be in treatment 2 and covariate K. In model (C) the effect of covariate K has been dropped, and the effect of this covariate is taken as the baseline

effect in the coefficients of the treatments.

$$\eta = \begin{cases} \sum_{j=1}^{2} \mu_j \delta_{nj} + \gamma \sum_{k=1}^{K} k\pi_{nk} & (A)\,, \\ \alpha + \mu\delta_{n1} + \sum_{k=1}^{K-1} \gamma_k \pi_{nk} & (B)\,, \\ \sum_{j=1}^{2} \mu_j \delta_{nj} + \sum_{j=1}^{2} \sum_{k=1}^{K-1} \gamma_{kj} \delta_{nj} \pi_{nk} & (C)\,. \end{cases}$$

In generalized linear models, in order to estimate the parameters, maximum likelihood estimation is used. Nevertheless, the specification of the likelihood function is not easy in many cases. If the link function is not correctly determined, it is imposible to give an explicit formula for the likelihood. Other common problem in order to determine the likelihood function is the dependence between observations, which could make very difficult its computation. This is the case of adaptive designs. A plausible alternative is the construction of the quasi-likelihood function, presented in [72]. For more details of the construction and characteristics of these function we refer to [48]. We can define the quasi-likelihood estimator, $\hat{\beta}_n$, as the solution of the equation

$$\sum_{i=1}^{n} g(\boldsymbol{\beta}^t \boldsymbol{x}_i)\boldsymbol{x}_i\{Y_i - \mu(\boldsymbol{\beta}^t \boldsymbol{x}_i)\} = 0\,,$$

where $g(\cdot)$ is usually $D\mu(\cdot)/v(\cdot)$. If the link function is the canonical one, $g(\cdot)$ becomes the identity, so the quasi-likelihood equation stands in this form,

$$\sum_{i=1}^{n} \boldsymbol{x}_i\{Y_i - \mu(\boldsymbol{\beta}^t \boldsymbol{x}_i)\} = 0\,.$$

Let $\lambda_{min}(A)$ and $\lambda_{min}(A)$ be, respectively, the minimum and maximum eigenvalue of the square matrix A. Then, the following technical lemma is introduced.

Lemma 5.2.1. *For models (A), (B), (C) of the lineal predictor, the next results are fulfilled:*

$$\lambda_{min}(D_n) \to \infty\,, \tag{5.4}$$

$$\liminf_{n\to\infty} \frac{\lambda_{min}(D_n)}{\lambda_{max}(D_n)^{1/2}(\log \lambda_{max}(D_n))^{1/2+\alpha}} \quad \textit{for some } \alpha > 0 \tag{5.5}$$

Proof. We compute the design matrix D_n and its determinant for each model.

Model (A): Denoting $F_n = \sum_{k=1}^{K} k\pi_{nk}$ and $\overline{F}_{nj} = \sum_{i=1}^{n} \delta_{ij} F_i$ for $j = 1, 2$ then,

$$D_n = \begin{pmatrix} N_{n.1} & 0 & \overline{F}_{n1} \\ 0 & N_{n.2} & \overline{F}_{n2} \\ \overline{F}_{n1} & \overline{F}_{n2} & \sum_{i=1}^{n} F_i^2 \end{pmatrix}$$

$$|D_n| = N_{n.1} N_{n.2} \left(\sum_{i=1}^{n} F_i^2 - N_{n.2} \overline{F}_{n1}^2 - N_{n.1} \overline{F}_{n2}^2 \right)$$

Model (B):

$$D_n = \begin{pmatrix} n & N_{n.1} & N_{n1.} & \dots & \dots & N_{nK-1.} \\ N_{n.1} & N_{n.1} & N_{n11} & \dots & \dots & N_{nK-11} \\ N_{n1.} & N_{n11} & N_{n1.} & 0 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & 0 & \ddots & \ddots & 0 \\ N_{nK-1.} & N_{nK-11} & 0 & \dots & 0 & N_{nK-1.} \end{pmatrix}$$

$$|D_n| = \prod_{i=1}^{K} N_{ni.} \left(N_{n.1} - \sum_{k=1}^{K} \frac{N_{nk1}^2}{N_{nk.}} \right)$$

Model (C):

$$D_n = \begin{pmatrix} N_{n.1} & 0 & N_{n11} & \dots & N_{nK-11} & 0 & 0 & 0 \\ 0 & N_{n.2} & 0 & 0 & 0 & N_{n12} & \dots & N_{nK-12} \\ N_{n11} & 0 & N_{n11} & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & 0 & \dots & \dots & 0 \\ N_{nK-11} & 0 & 0 & 0 & N_{nK-11} & 0 & \dots & 0 \\ 0 & N_{n12} & 0 & \dots & 0 & N_{n12} & 0 & 0 \\ \vdots & \vdots & 0 & \dots & \dots & 0 & \ddots & 0 \\ 0 & N_{nK-12} & 0 & \dots & \dots & \dots & 0 & N_{nK-12} \end{pmatrix}$$

$$|D_n| = \prod_{i=1}^{K} N_{nk1} N_{nk2}$$

For model (A), the lemma is proven in Lemma 3.1 in [53]. For models (B) and (C), all the elements of the $D_n = \{d_{ij}(n)\}$ matrix are $\Theta(n)$ functions, so, we can ensure the irreducibility of the matrix and apply Theorem 1.5 in [68], which guarantees the next inequality,

$$\min_i \left(\sum_{j=i}^{l} d_{ij}(n) \right) \leq \lambda_{max}(D_n) \leq \max_i \left(\sum_{j=i}^{l} d_{ij}(n) \right).$$

It follows that $\lambda_{max}(D_n) \in \Theta(n)$. Besides, $|D_n| \in \Theta(n^l)$ in models (B) and (C), with $l = dim(D_n)$. Using

$$(\lambda_{min}(D_n))^{l-1} \lambda_{max}(D_n) \leq |D_n| \leq (\lambda_{max}(D_n))^{l-1} \lambda_{min}(D_n),$$

it holds that $\lambda_{min}(D_n) \in \Theta(n)$ and (5.4) and (5.5) follow straightforwardly.

□

In the next proposition, the properties of the estimators $\hat{\beta}_n$ are studied under a general stratified design and the canonical link function.

Proposition 5.2.1. *Assume that n patients have been allocated in the treatments with a stratified design. Let N_{nkj} the total number of patients, up to the n-th patient, in the stratum k allocated in treatment j, with $k = 1, ..., K$ and $j = 1, 2$.*

Assume that patients' responses $\{Y_n\}$ satisfy a generalized linear model (5.3) where μ is the logistic function and $\boldsymbol{\beta}$ is as in model (A), (B) or (C).

Assume also that for each pair k and j, there exists a vector $\mathbf{u}_k = (u_{k1}, u_{k2})$, with $u_{k1} > 0$ and $u_{k2} > 0$ such that

$$N_{nkj}/n \to u_{kj}, \quad k = 1, \ldots, K, \quad j = 1, 2. \tag{5.6}$$

Let $\hat{\boldsymbol{\beta}}_n$ be the MQLE of the vector of parameters $\boldsymbol{\beta}$, then, as $n \to \infty$

$$\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}, \ \ a.s.$$

$$V_n^{-C/2} V_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^t \to N(0, I) \quad [D]\,,$$

with

$$V_n = \sum_{i=1}^{n} \mathbf{x}_i^t H(\boldsymbol{\beta}^t \boldsymbol{x}_i))\boldsymbol{x}_i$$

where $H(t)$ is the derivative of μ evaluated in $t \in \mathbb{R}$ and $V_n^{C/2}$ is the lower triangular matrix of the Cholesky square root.

Proof. The result follows if conditions in Theorem 2 in [76] hold.

$A1$) For any $t \in \mathbb{R}$, $\Sigma^{-1} > 0$, $\det H(t) \neq 0$, each element of $\Sigma^{-1} > 0$ is continuously differentiable and $\mu(t)$ is twice continuously derivable.

$A2^*$) With probability 1, $\sup_{i\geq 1} \|\boldsymbol{x}_i\| < \infty$, $\lambda_{min}(D_n) \to \infty$ and

$\liminf_{n\to\infty} \frac{\lambda_{min}(D_n)}{\lambda_{max}(D_n)^{1/2}(\log \lambda_{max}(D_n))^{1/2+\alpha}}$ for some $\alpha > 0$.

$A3^*$) For each $i \geq 1$, y_i is $\mathcal{F}_i$ measurable, $\boldsymbol{x}_i$ is $\mathcal{F}_{i-1}$ measurable. With probability 1, $E(y_i \mid \mathcal{F}_{i-1}) \equiv \mu(\boldsymbol{\beta}^t \boldsymbol{x}_i)$ and $\sup_{i\geq 1} E(\|\boldsymbol{\epsilon}_i\| \mid \mathcal{F}_{i-1}) < \infty$

$A4^*$) With probability 1, $Cov(\epsilon_i|\mathcal{F}_{i-1}) > cI$ and $\sup_{i\geq 1} E(\|\boldsymbol{\epsilon}_i\|^r \mid \mathcal{F}_{i-1})$ for some $r > 2$. Besides, there exists a non-random positive definite symmetric matrix F_n such that

$$F_n^{\frac{1}{2}} V_n F_n^{\frac{1}{2}} \to I \quad \text{in probability.}$$

Conditions $A1$ and $A3^*$ are immediate to check for (5.3) under the three models, due to the definition of the logistic model and the boundedness of the estimators. Condition $A2^*$ is a direct conclusion of Lemma 5.2.1.

For condition $A4^*$, from the definition of the error terms, $Cov(\epsilon_i|\mathcal{F}_{i-1}) > cI$ and $\sup_{i\geq 1} E\left[\|\epsilon_i\|^r|\mathcal{F}_{i-1}\right] < \infty$ are direct. As $V_n = \sum_{i=1}^n v_i \boldsymbol{x}_i \boldsymbol{x}_i^t$, being $v_i = H(\boldsymbol{\beta}^t \boldsymbol{x}_i) = \frac{e^{\beta^t \boldsymbol{x}_i}}{(e^{\beta^t \boldsymbol{x}_i}+1)^2}$, V_n is symmetric and definite positive. If $V_n/n \to C$ a.s. when $n \to \infty$ and $\|C\| \neq 0$, then C is also symmetric and definite positive. So, there exits as symmetric and positive matrix F such that $C = F^{1/2}F^{1/2}$. Defining $F_n = nF$, we have that $F_n^{-1/2} V_n F_n^{-1/2} \to I$ as $n \to \infty$.

We prove now that $V_n/n \to C$ a.s. when $n \to \infty$ and $\|C\| \neq 0$. Observe that once δ_{i1} and F_i are known, v_i does not depend on i and we denote as v_{kj} the value of v_i if the i-th

patient has prognostic factor k and is allocated to treatment j. Under model (A) we have that

$$V_n = \begin{pmatrix} \sum_{i=1}^{n} v_i\delta_{i1} & 0 & \sum_{i=1}^{n} v_i\delta_{i1}F_i \\ 0 & \sum_{i=1}^{n} v_i\delta_{i2} & \sum_{i=1}^{n} v_i\delta_{i2}F_i \\ \sum_{i=1}^{n} v_i\delta_{i1}F_i & \sum_{i=1}^{n} v_i\delta_{i2}F_i & \sum_{i=1}^{n} v_iF_i^2 \end{pmatrix}. \tag{5.7}$$

The analysis of the convergence is particular for each element of the matrix. We illustrate the procedure with $\sum_{i=1}^{n} v_i\delta_{i1}F_i$ and the convergence of the other elements will follow with the same reasoning. Knowing δ_{i1} and π_{ik}, v_i is V_{k1}, so

$$\sum_{i=1}^{n} v_i\delta_{i1}F_i = \sum_{i=1}^{n}\sum_{k=1}^{K} v_i\delta_{i1}k\pi_{ik} = \sum_{k=1}^{K} v_{k1}N_{nk1}\,. \tag{5.8}$$

Since $N_{nk1}/n \to u_{k1} > 0$, $\frac{1}{n}\sum_{i=1}^{n} v_i\delta_{i1}F_i$ also converges.

Reasoning in the same way, the results also follow under model (B) and model (C). □

Corollary 5.2.1. *If the stratified Klein urn design has been applied, under models (A), (B) and (C), the MQL estimators $\hat{\boldsymbol{\beta}}_n$ are strongly consistent and asymptotically normal.*

Proof. Under the stratified Klein urn design, using Proposition 5.1.1, the convergence $N_{nij}/n \to u_{ij}$ is ensured, and then Proposition 5.2.1 follows directly. □

This fact permits the use of these estimators in the inferential study via simulation in the next section.

5.3 Simulation Study

The previous section was focused on the theoretical analysis of the properties of the MQL estimators in a generalized linear model. Now, an inferential comparative study based on simulation is presented, in order to test the hypothesis of equality between both treatments. The comparative study is made among the designs studied in Chapter 3, complete randomization, Efron's design, PTW design, Klein urn design, DBCD design with $\alpha = 2$

and ERADE design with $\alpha = 7$. We are going to follow two different approaches. On the one hand, a population model is assumed with dichotomous responses. Then, a generalized linear model, more precisely a logistic model, is considered for these responses, as in the previous section. Besides, a time trend factor is introduced, defining a covariate F_n with three levels, values $1, 2$ or 3. So the model stands as

$$logit(y_n) = \mu_1\delta_{n1} + \mu_2(1 - \delta_{n2}) + \gamma F_n \tag{5.9}$$

which corresponds with model (A) of the previous sections. Our target is to estimate the difference of the effect of the treatments,

$$H_0 : \mu_1 - \mu_2 = 0, \qquad H_1 : \mu_1 - \mu_2 \neq 0.$$

The test statistic is defined as $T_{glm} = (1, -1, \mathbf{0})\hat{\boldsymbol{\beta}} = \hat{\mu}_1 - \hat{\mu}_2$. To ensure the asymptotic normality of these estimators, note that under a population model, all the designs considered satisfy that, almost surely, $N_{n.i}/n \to \rho$, where $0 < \rho < 1$ varies depending on the design. This result is (5.6) condition .

On the other hand, assuming a randomization model, RBI . The procedure is the same as the one presented in Chapter 3. The null hypothesis is that both treatments behave equally. The test statistic is $T_n = S_{n,1}/N_{n,1} - S_{n,2}/N_{n,2}$, which is the difference of the proportion of successes in each treatment. This statistic can be seen as a randomization version of the classical test of difference of means.

In Table 5.1, $\mu_2 = 0.1$, $\beta = 1$ and the following two situations: $\mu_1 = 0.1$, when there is not difference between treatments and the null hypothesis is true, and $\mu_1 = 2.1$, when the null hypothesis is false. The covariate $\{F_n\}$ is simulated as a sequence of independent and identically distributed uniform random variables on the set $\{1, 2, 3\}$. We consider $n = 50, 100, 200$ patients. The responses are simulated using the GLM for both cases. We make asymptotic inference for the GLM estimators following [76] and the RBI as in Table 3.5.

Table 5.1: Proportion of rejections with the randomization test and test statistic T (Monte Carlo estimations of the p-values with 2,500 runs) and the classical parametric inference with the MQL estimators of the GLM.

		$n = 50$		$n = 100$		$n = 200$	
μ_1, μ_2, β	D	T	GLM	T	GLM	T	GLM
0.1, 0.1, 1	CR	0.044	0.019	0.037	0.032	0.052	0.046
	Efron	0.040	0.016	0.038	0.034	0.0548	0.050
	PTW	0.061	0.041	0.053	0.052	0.038	0.39
	Klein	0.040	0.020	0.062	0.060	0.049	0.052
	DBCD2	0.036	0.017	0.056	0.049	0.056	0.048
	ERADE7	0.042	0.026	0.049	0.051	0.060	0.051
2.1, 0.1, 1	CR	0.263	0.019	0.580	0.230	0.904	0.760
	Efron	0.198	0.018	0.520	0.217	0.867	0.763
	PTW	0.221	0.047	0.480	0.288	0.852	0.798
	Klein	0.228	0.058	0.559	0.260	0.825	0.758
	DBCD2	0.235	0.025	0.587	0.258	0.877	0.747
	ERADE7	0.210	0.034	0.561	0.289	0.863	0.767

Observe that the randomization test performs better than the standard test for parameters in a GLM model with $n = 50$ and $n = 100$ sample sizes. For $n = 200$ results are quite similar for both inferential procedures. We do not appreciate a different behavior depending on the design applied. The bad figures for the GLM estimation can be explained because, as stated before, they are based on the asymptotic normality of the estimators and higher values of n are needed. Nevertheless, for larger sample sizes, as $n = 200$, the results improve because conditions in [28] hold for this model and asymptotic normality is true even if an adaptive design has been applied for randomization. So that, RBI seems a fruitful inference procedure when tendency can be expected in the responses and randomization with adaptive designs is applied. Observe that in order to obtain the distribution of T_n nor a population model, neither an statistical model are assumed. Besides this, a simulation study shows that for small to moderate sample sizes the inference

based on T_n is more powerful than the inference based on T_{glm} even if the population model is appropriate for model (A) and the values assumed for success probabilities in the simulations.

Conclusions and future work

In the framework of a probabilistic analysis of adaptive experimental design, this thesis focuses on the description and analysis of a new response-adaptive design: the Klein urn design. On the one hand, the asymptotic structure of the Klein urn design has been analyzed and a parametric inferential study has been developed. The behavior of the Klein urn design is asymptotically very similar to the behavior of the drop-the-loser design, which is considered as a very competitive design in the literature. On the other hand, a thorough study of the exact properties of the Klein urn design can be made for any sample size n due to its probabilistic structure, based on Markov Chains. This is an advantage with respect to the drop-the-loser design for which exact properties are not well studied.

Along with the parametric inference, randomization based inference has also been studied for response-adaptive designs. The use of adaptive designs complicates the randomization based inference process, because all the allocation arrangements are not equally probable as in complete randomization, but these difficulties have been overcome and an exact and asymptotic study has been completed. Recurrence relations have been used to obtain the exact distributions of some statistics and the p-values associated to these statistics. In addition central limit theorems for these test statistics have been proven. This asymptotic result gives an alternative to the use of Monte Carlo approximation which is more familiar in clinical research practice.

The behavior of the Klein urn design has been analyzed when covariates are included in the design, extending the spectrum of its applicability. Ad-hoc estimators under a stratified design and also the maximum quasi likelihood estimators of the generalized linear model have been studied and their properties have been obtained. Finally, a simulation study has been carried out comparing randomization based inference and inference based on generalized linear models. It has been concluded that the RBI has not a worse behavior than inference based on GLM, thereby it becomes a viable alternative, with the advantage that is free of model specifications.

Some results can be extended or generalized. In Chapter 2 there is a conjecture about the limit of the maximum eigenvalue, which is related to a measurement of the accidental bias. In Theorem 4.2.1 milder conditions on the coefficients could be of interest. A study of the speed of convergence would complete the different limit laws obtained. Dichotomous responses have been arisen in this thesis, so a generalization to other kind of responses would be of interest, specially in the algorithm for obtaining the exact distribution of the test statistic S_n. It would also be of interest a randomization based inferential study for other statistics as, for instance, the pooled statistic presented in Remark 5.1.1. and the inference for different link functions when generalized linear models are used.

Conclusiones y trabajo futuro

Esta tesis se ha centrado en la presentación y estudio del diseño de urna de Klein. Por un lado, se ha analizado su estructura asintótica y se ha hecho un estudio inferencial paramétrico, siendo el comportamiento asintótico del diseño de urna de Klein muy parecido al diseño drop-the-loser, reconocido en la literatura por sus buenas propiedades. Por otro lado, por su estructura probabilística basada en cadenas de Markov, se tiene un conocimiento amplio sobre sus propiedades exactas, para cualquier tamaño de muestra, mejor que el disponible para el diseño DTL.

Además de la inferencia paramétrica, se ha profundizado en la inferencia no paramétrica basada en la aleatorización. La utilización de diseños adaptativos presenta algunas dificultades para realizar la inferencia basada en la aleatorización: en general, no todas las configuraciones o reordenaciones de tratamientos son factibles, y las configuraciones posibles pueden tener distintas probabilidades. Se han resuelto estas dificultades y se han obtenido resultados exactos y asintóticos. Por una parte, se han obtenido resultados de recurrencia que permiten calcular la distribución exacta de diversos estadísticos y de los p-valores para los contrastes asociados a esos estadísticos. Estos resultados son aplicables para tamaños muestrales pequeños o moderados. Por otra parte, se han obtenido teoremas centrales del límite para algunos de estos estadísticos. Estos resultados proporcionan una alternativa sencilla a la utilización de técnicas de aproximación de Monte Carlo. Además, esta aproximación por una distribución normal es más cercana a la práctica habitual de los investigadores clínicos.

Finalmente, también se ha analizado el comportamiento del diseño de urna de Klein ante la presencia de covariables, ampliando el espectro de su aplicabilidad. Se han demostrado las propiedades de estimadores ad-hoc en el caso estratificado y de los estimadores de máxima cuasi-verosimilitud para el modelo lineal generalizado, haciendo posible un estudio de inferencia también en este caso. Se ha comparado la inferencia basada en la aleatorización y la basada en estimadores de los modelos lineales generalizados, y se

ha concluido que a pesar de requerir menos hipótesis, la inferencia basada en la aleatorización no tiene un comportamiento peor, con lo que se convierte en una alternativa válida y factible, con la ventaja de estar libre de especificaciones.

Algunos resultados son ampliables o generalizables. En el capítulo 2 se ha enunciado una conjetura acerca del comportamiento asintótico de un valor propio máximo, relacionado con el cálculo del sesgo accidental. En el Teorema 4.2.1, las condiciones sobre los coeficientes son exigentes, así que una relajación de éstas sería de interés. En los diferentes teoremas límite demostrados, sería interesante también hacer un estudio de velocidades de convergencia. En la tesis se han considerado respuestas dicotómicas. La generalización a otro tipo de respuestas es un campo abierto, siendo de especial interés el cálculo del estadístico S_n para cualquier tipo de respuesta. En el capítulo 5, sería interesante un estudio inferencial basado en la aleatorización para el estadístico ponderado del Remark 5.1.1. y también se podrían considerar diferentes funciones de enlace en los modelos lineales generalizados, asociados a diferentes tipos de respuestas.

La realización de esta tesis ha sido posible gracias a una beca predoctoral del Gobierno de Navarra.

Bibliography

[1] Aickin, M. Randomization, balance, and the validity and efficiency of design-adaptive allocation methods. *Journal of Statistical Planning and Inference 94*, 1 (2001), 97–119.

[2] Atkinson, A. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika 69*, 1 (1982), 61–67.

[3] Atkinson, A. C., and Biswas, A. *Randomised response-adaptive designs in clinical trials*, vol. 764. Chapman and Hall/CRC, Boca Raton, 2014.

[4] Balakrishnan, N., and Koutras, M. V. *Runs and scans with applications*, vol. 764. John Wiley & Sons, New York, 2011.

[5] Baldi Antognini, A. On the speed of convergence of some urn designs for the balanced allocation of two treatments. *Metrika 62*, 2-3 (2005), 309–322.

[6] Baldi Antognini, A., and Giannerini, S. Generalized Pólya urn designs with null balance. *Journal of applied probability 44*, 3 (2007), 661–669.

[7] Baldi Antognini, A., and Giovagnoli, A. A new biased coin design for the sequential allocation of two treatments. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 53*, 4 (2004), 651–664.

[8] Basu, D. Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association 75*, 371 (1980), 575–582.

[9] Begg, C. B. On inferences from Wei's biased coin design for clinical trials. *Biometrika 77*, 3 (1990), 467–473.

[10] Begg, C. B., and Iglewicz, B. A treatment allocation procedure for sequential clinical trials. *Biometrics* (1980), 81–90.

[11] Bélisle, C., and Melfi, V. Independence after adaptive allocation. *Statistics & Probability Letters 78*, 3 (2008), 214–224.

[12] Chen, Y.-P. Which design is better? Ehrenfest urn versus biased coin. *Advances in Applied Probability* (2000), 738–749.

[13] Chow, S.-C., and Chang, M. *Adaptive design methods in clinical trials*. CRC Press, Boca Raton, 2010.

[14] Cinlar, E. *Introduction to stochastic processes*. Prentice-Hall, Englewood Cliffs, NJ, 1975.

[15] Cook, R. D., and Wong, W. K. On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association 89*, 426 (1994), 687–692.

[16] Cox, D. A remark on randomization in clinical trials. *Utilitas Mathematica A 21* (1982), 245–252.

[17] DasGupta, A. *Asymptotic theory of statistics and probability*. Springer, New York, 2008.

[18] Edgington, E. S. *Randomization tests*. Marcel Dekker, New York, 1987.

[19] Efron, B. Forcing a sequential experiment to be balanced. *Biometrika 58*, 3 (1971), 403–417.

[20] Eisele, J. R. The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference 38*, 2 (1994), 249–261.

[21] Eisele, J. R., and Woodroofe, M. B. Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics* (1995), 234–254.

[22] Feller, W. *Introducción a la teoría de probabilidades y sus aplicaciones*, vol. 1. Limusa Méxicoˆ eD. FDF, 1978.

[23] Flournoy, N., Galbete, A., Moler, J. A., and Plo, F. Randomization based inference for the drop-the-loser rule. In *mODa 10–Advances in Model-Oriented Design and Analysis*. Springer, 2013, pp. 81–89.

[24] Flournoy, N., Haines, L. M., and Rosenberger, W. F. A graphical comparison of response-adaptive randomization procedures. *Statistics in Biopharmaceutical Research 5*, 2 (2013), 126–141.

[25] Galbete, A., Moler, J. A., and Plo, F. Asymptotic behavior of randomization tests for the Klein urn design. *Annals of the Institute of Statistical Mathematics*, Submitted.

[26] Galbete, A., Moler, J. A., and Plo, F. A response-driven adaptive design based on the Klein urn. *Methodology and Computing in Applied Probability*, DOI 10.1007/s11009–013–9344–9.

[27] Galbete, A., Moler, J. A., and Plo, F. Response-adaptive designs based on the Ehrenfest urn. *Monografías matemáticas Garcia Galdeano 38* (2013), 85–96.

[28] Gao, Q.-B., Lin, J.-G., Zhu, C.-H., and Wu, Y.-H. Asymptotic properties of maximum quasi-likelihood estimators in generalized linear models with adaptive designs. *Statistics 46*, 6 (2012), 833–846.

[29] Good, P. *Permutation, parametric and bootstrap tests of hypotheses*. Springer, New York, 1994.

[30] Hall, P., and Heyde, C. C. *Martingale limit theory and its application*. Academic press, New York, 1980.

[31] Hollander, M., and Peña, E. Nonparametric tests under restricted treatment-assignment rules. *Journal of the American Statistical Association 83*, 404 (1988), 1144–1151.

[32] Hu, F., and Rosenberger, W. F. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association 98*, 463 (2003), 671–678.

[33] Hu, F., and Rosenberger, W. F. *The theory of response-adaptive randomization in clinical trials*, vol. 525. Wiley, New York, 2006.

[34] Hu, F., Rosenberger, W. F., and Zhang, L.-X. Asymptotically best response-adaptive randomization procedures. *Journal of Statistical Planning and Inference 136*, 6 (2006), 1911–1922.

[35] Hu, F., and Zhang, L.-X. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics 32*, 1 (2004), 268–301.

[36] Hu, F., Zhang, L.-X., and He, X. Efficient randomized-adaptive designs. *The Annals of Statistics* (2009), 2543–2560.

[37] Hu, Y., and Hu, F. Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics 40*, 3 (2012), 1794–1815.

[38] Isaacson, D. L., and Madsen, R. W. *Markov chains, theory and applications.* Wiley, New York, 1976.

[39] Ivanova, A. A play-the-winner-type urn design with reduced variability. *Metrika 58*, 1 (2003), 1–13.

[40] Katti, S. Moments of the absolute difference and absolute deviation of discrete distributions. *The Annals of Mathematical Statistics 31*, 1 (1960), 78–85.

[41] Klein, M. J. Generalization of the Ehrenfest urn model. *Physical Review 103*, 1 (1956), 17.

[42] Knopp, K. *Theory and application of infinite series*. Courier Dover Publications, New York, 1951.

[43] Lai, T. L., and Wei, C. Z. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics 10*, 1 (1982), 154–166.

[44] Liu, G., and Liu, W. On the strong law of large numbers for functionals of countable nonhomogeneous markov chains. *Stochastic processes and their applications 50*, 2 (1994), 375–391.

[45] Mahmoud, H. *Pólya urn models*. CRC press, Boca Raton, 2008.

[46] Markaryan, T., and Rosenberger, W. F. Exact properties of Efrons biased coin randomization procedure. *The Annals of Statistics 38*, 3 (2010), 1546–1567.

[47] May, C., and Flournoy, N. Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *The Annals of Statistics* (2009), 1058–1078.

[48] McCullagh, P., and Nelder, J. A. *Generalized linear model*, vol. 37. Chapman & Hall/CRC, London, 1989.

[49] Mehta, C. R., Patel, N. R., and Senchaudhuri, P. Efficient monte carlo methods for conditional logistic regression. *Journal of The American Statistical Association 95*, 449 (2000), 99–108.

[50] Mehta, C. R., Patel, N. R., and Wei, L. Constructing exact significance tests with restricted randomization rules. *Biometrika 75*, 2 (1988), 295–302.

[51] Melfi, V. F., and Page, C. Estimation after adaptive allocation. *Journal of statistical planning and inference 87*, 2 (2000), 353–363.

[52] Melfi, V. F., Page, C., and Geraldes, M. An adaptive randomized design with application to estimation. *Canadian Journal of Statistics 29*, 1 (2001), 107–116.

[53] Moler, J. A., Plo, F., and San Miguel, M. A sequential design for a clinical trial with a linear prognostic factor. *Journal of Statistical Planning and Inference 137*, 12 (2007), 3964–3974.

[54] Nelder, J. A., and Wedderburn, R. W. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)* (1972), 370–384.

[55] Peligrad, M. Central limit theorem for triangular arrays of non-homogeneous markov chains. *Probability Theory and Related Fields 154*, 3-4 (2012), 409–428.

[56] Pesarin, F., and Salmaso, L. *Permutation tests for complex data: theory, applications and software*. Wiley, Chichester, 2010.

[57] Plamadeala, V., and Rosenberger, W. F. Sequential monitoring with conditional randomization tests. *The Annals of Statistics 40*, 1 (2012), 30–44.

[58] Pocock, S. J. *Clinical trials: a practical approach*. John Wiley & Sons, Chichester, 1984.

[59] Pocock, S. J., and Simon, R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* (1975), 103–115.

[60] Rosenberger, W. F. Asymptotic inference with response-adaptive treatment allocation designs. *The Annals of Statistics* (1993), 2098–2107.

[61] Rosenberger, W. F., Flournoy, N., and Durham, S. D. Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *Journal of Statistical Planning and Inference 60*, 1 (1997), 69–76.

[62] Rosenberger, W. F., and Hu, F. Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials 1*, 2 (2004), 141–147.

[63] Rosenberger, W. F., and Lachin, J. M. *Randomization in clinical trials: theory and practice*. Wiley, New York, 2002.

[64] Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N., and Ricks, M. L. Optimal adaptive designs for binary response trials. *Biometrics 57*, 3 (2001), 909–913.

[65] Rosenberger, W. F., and Sverdlov, O. Handling covariates in the design of clinical trials. *Statistical Science 23*, 3 (2008), 404–419.

[66] Rosenberger, W. F., Vidyashankar, A., and Agarwal, D. K. Covariate-adjusted response-adaptive designs for binary response. *Journal of biopharmaceutical statistics 11*, 4 (2001), 227–236.

[67] Rosenberger William, F. *Asymptotic inference problems arising from clinical trials using response-adaptive treatment allocation*. PhD thesis, George Washington University, 1992.

[68] Seneta, E. *Non-negative Matrices and Markov Chains*. Springer Verlag, New York, 1981.

[69] Shao, J., Yu, X., and Zhong, B. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika 97*, 2 (2010), 347–360.

[70] Smith, R. L. Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society. Series B (Methodological)* (1984), 519–543.

[71] Smythe, R., and Wei, L. Significance tests with restricted randomization design. *Biometrika 70*, 2 (1983), 496–500.

[72] Wedderburn, R. W. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika 61*, 3 (1974), 439–447.

[73] Wei, L. An application of an urn model to the design of sequential controlled clinical trials. *Journal of the American Statistical Association 73*, 363 (1978), 559–563.

[74] Wei, L. Exact two-sample permutation tests based on the randomized play-the-winner rule. *Biometrika 75*, 3 (1988), 603–606.

[75] Wei, L., and Durham, S. The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association 73*, 364 (1978), 840–843.

[76] Yin, C., Zhao, L., and Wei, C. Asymptotic normality and strong consistency of maximum quasi-likelihood estimates in generalized linear models. *Science in China Series A 49*, 2 (2006), 145–157.

[77] Zhang, L., Chan, W. S., Cheung, S. H., and Hu, F. A generalized drop-the-loser urn for clinical trials with delayed responses. *Statistica Sinica 17*, 1 (2007), 387.

[78] Zhang, L.-X., Hu, F., Cheung, S. H., and Chan, W. S. Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics 35*, 3 (2007), 1166–1182.

[79] Zhang, L.-X., Hu, F., Cheung, S. H., and Chan, W. S. Immigrated urn modelstheoretical properties and applications. *The Annals of Statistics 39*, 1 (2011), 643–671.