ADAPTIVE DESIGN TOWARDS OPTIMAL DOSE

ZHANG JIANCHUN

NATIONAL UNIVERSITY OF SINGAPORE

2004

ADAPTIVE DESIGN TOWARDS OPTIMAL DOSE

ZHANG JIANCHUN

(B.Sc. Univ. of Science & Technology of China)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2004

To Wenjia

Acknowledgements

For the completion of this master thesis, I would like to take this opportunity to express my deepest gratitude to my supervisor Prof. Bai, Zhidong for all his invaluable advice and guidance, endless patience and encouragement during my two years' study in NUS. Without his numerous valuable suggestions and comments and his generous help, this assay would be impossible to be completed. I would also like to thank you for the faculty and the graduate students, who teach me and help me a lot during my studies.

Specifically, I am really grateful to Miss Li, Wenjia, for her love and care, for her understanding and sustaining support in spirit of these two years' study and the incoming PhD study. I would also give my sincere thanks to my family for their understandings, especially, my mother.

Contents

List of Figures

List of Tables

Summary

Adaptive designs have been proposed for ethical concerns, their characteristics, especially in statistics, are widely investigated in the literature. In this thesis, we investigate adaptive designs which direct the trials to an optimal dose level by using Generalized Friedman's Urn Model (GFU), considering the patients' effect on the probability of success for each dose. A generalized linear model(GLM) is employed with consideration of the patients' covariates and the dose levels simultaneously. The limiting properties of the Maximum Likelihood Estimation(MLE), especially its Central Limit Theorem (CLT) are established in the circumstance that the response variables are dependent. The asymptotic properties of Urn composition and allocation proportion are investigated. Simulations are conducted to verify these properties.

Key Words: Adaptive Design; Generalized Friedman's Urn Model; Generalized Linear Model; Maximum Likelihood Estimation; Urn Composition

Chapter 1

Introduction

1.1 Adaptive Designs

In any sequential medical experiment on a cohort of human beings, there is an ethical imperative to provide the best possible medical care for the individual patient. This ethical imperative may be compromised if a traditional randomization scheme involving 50-50 allocation is used as accruing evidence to favor one treatment over the other.

A case in point is reported by Conner et al. (1994) to evaluate the hypothesis that the antiviral therapy AZT reduced the risk of maternal-to-infant HIV transmission. A traditional randomization scheme was used to obtain equal allocation to both AZT and placebo, resulting in 239 pregnant women receiving AZT and 238

receiving placebo. The endpoint was whether the newborn infant was HIV-negative or HIV-positive. The results of the trial were compelling: at the conclusion of the trial, 60 newborns were HIV-positive in the placebo group and only 20 newborns were HIV-positive in the AZT group. Three times as many infants on placebo have received a death sentence by the transmission of HIV from their mothers. If they had been given AZT, one could presume that many would have been saved. Given the seriousness of the outcome of this study, it is reasonable to argue that 50-50 allocation was unethical. As accruing information favoring the AZT arm became available, allocation probabilities should have been shifted from 50-50 allocation proportional to the weight of evidence for AZT. Designs which attempt to do this are called *adaptive designs, response-adaptive designs or response-driven designs.*

Adaptive design in clinical trials are schemes for patient allocation to treatment, the goal of which is to place more patients on the better treatment based on patient responses already accrued in the trials. For example, if there are two treatments A and B, then when a patient is ready to be allocated to a treatment, if treatment A appears to be more successful than treatment B, that patient would have a greater than 50% chance of being allocated to treatment A. Adaptive designs are attractive because they satisfy an ethical imperative of caring for the individual patient in a group experiment, while allowing for the same group inference.

In statistics, sequential design is a subfield of experimental design which deals with the appropriate sequential selection of design points. When design points are

sequentially selected according to outcomes at previously selected design points, such designs are called *adaptive*. Since future design point selection can rely on information previously accrued, they can target an objective more accurately than if design points are selected in the absence of information.

1.2 Urn Model

In adaptive designs, the allocation rules of patients in the trials are primary concerns. The ethics of clinical trials not only need to derive information about the effectiveness of the treatments, but to benefit the health of patients as well. Urn models have been one technique(among many) used to incorporate accruing data into the sequential design.

1.2.1 Play-the-Winner Rule

From the perspective of ethics, Zelen (1969) firstly explored the design methods to place more patients on the better treatment and proposed out the original design called *Play-the-Winner Rule*. From then on, allocating patients sequentially in clinical trials has been extensively explored in theoretical fields. In Zelen's formulation there are two treatments (say, A and B), patients enter the trial sequentially and are allocated to treatment A or B, and the trial outcome is success

or failure depends only on the treatment given. A success on a particular treatment generated a future trial on the same treatment assigned to the next patient. A failure on a treatment generates a future trial on the alternate treatment. When there exists delayed responses, i.e. the result of the treatment can not be obtained until the next patient enters the trials, allocation is determined by tossing a fair coin. In Play-the-Winner Rule, the allocation scheme is not random but deterministic, which may bias the trial in various ways. For example, if the experimenter is in favor of treatment A and he knows or guesses which treatment will be the next assignment, then he may introduce bias into the trial through the selection of patients, this kind of bias is called selection bias. On the other hand, it does not take the case of the delayed responses into consideration. However, this design can be seen as a raw urn model implicitly and widened the researchers' horizon to the randomized urn models later on.

1.2.2 Randomized Play-the-Winner Rule

Wei and Durham (1978) propose the *Randomized Play-the-Winner* (RPW) rule which keeps the spirit of the *Play-the-winner* rule in that it assigns more patients to the better treatment. But this rule has the advantages that it is not deterministic and is less vulnerable to experimental bias, it allows delayed response by the patients. The formulation of the RPW rule is as follows: Assume there are two treatments(say, A and B), with dichotomous response(success or failure). We

start with u balls of each type in the urn. When a patient is ready for randomization, a ball is drawn and replaced, and the appropriate treatment is assigned. If the response of the patient is a success, an additional β balls of the same type are added to the urn and an additional α balls of the opposite type are added to the urn. If the response was a failure, then an additional α balls of the same type are added to the urn and additional β balls of the opposite type are added to the urn. Wei and Durham compared the RPW with the traditional equal-allocation rule and found that the latter can involve selection bias. The above RPW can be denoted as RPW(u, α, β). Compared to the Play-the-Winner rule, RPW is a true randomized urn model, which allows the delayed responses and take advantage of randomization strategy.

1.2.3 Generalized Friedman's Urn Model

A very important class of adaptive designs is one based on the generalized Friedman's urn (GFU) model (Athreya and Karlin (1968)), which has wide applications in clinical trials, bioassay and psychophysics.

Adaptive designs using the GFU model can be formulated as follows. Assume, initially, an urn contains K types of balls, denoted by $Y_0 = (Y_{01}, ...Y_{0K})$, respectively representing K treatments in a clinical trial, where Y_{0k} denotes the number of balls of type $k, k = 1, ..., K$. These treatments are to be sequentially

assigned in n consecutive stages. At stage $i, i = 1, ..., n$, a ball is drawn from the urn with replacement. If a ball of type k is drawn at the i-th stage, then the treatment k is assigned to the patient $i, k = 1, ..., K; i = 1, ..., n$. Let $\xi(i)$ denote a random variable associated with the i-th stage of the clinical trial, which may include measurements on the i-th patient and the outcome of the treatment at the i-th stage. Then additional $D_{k,\,q}(i)$ balls of type q are added to the urn, $q = 1, ..., K$, where $D_{k,\,q}(i)$ is a function of $\xi(i)$. This procedure is repeated to the n-th stage. After n stages and generations, the urn composition is denoted by the vector $Y_n = (Y_{n1}, ..., Y_{nK})$, where Y_{nk} represents the number of type k balls in the urn. Furthermore, we define $D_i = \langle\langle D_{k,\,q}(i), k, q = 1, ..., K\rangle\rangle$ and $H_i = \langle\langle E(D_{k,\,q}(i)), k, q = 1, ..., K\rangle\rangle, i = 1, ..., n.$(sometimes, the conditional expectation). The matrices D_i's are called adding rules and H_i's are the *generating matrices*.

We call the GFU model *homogeneous* if $H_i = H$ for all $i = 1, ..., n$. For a homogeneous GFU model, under the assumptions (i) $Pr\{D_{k,\,q} = 0, q = 1, ..., K\} = 0$ for every $k = 1, ..., K$ and H is positive regular(i.e. H^m has positive entries for some $m > 0$),Athreya and Karlin (1968) and Athreya and Ney (1972) prove the following results for the Generalized Friedman's Urn model:

$$\frac{N_{nk}}{n} \xrightarrow{a.s.} v_k \tag{1.1}$$

$$\frac{Y_{nk}}{\sum_{q=1}^{K} Y_{nq}} \xrightarrow{a.s.} v_k \tag{1.2}$$

where N_{nk} is the number of patients allocated to the k-th treatment after n stages.

$v = (v_1, ...v_K)$ is the left eigenvector of the largest eigenvalue λ of H. Let λ_1 denote the eigenvalue of the second largest real part, with corresponding right eigenvector η. Furthermore, under additional assumption that $\lambda > 2\mathrm{Re}(\lambda_1)$, Athreya and Karlin (1968) show that

$$n^{-\frac{1}{2}} Y_n \eta' \to N(0, \sigma^2) \tag{1.3}$$

where σ^2 is constant.

Wei (1979) first noted that the RPW rule could be formulated as a GFU model. Let p_i be the probability of success on treatment $i = A, B$,and $q_i = 1 - p_i$. The RPW rule is a Generalized Friedman's Urn with $K = 2$; the adding rule is given by:

$$D = \begin{bmatrix} \xi_A\beta + (1-\xi_A)\alpha & \xi_A\alpha + (1-\xi_A)\beta \\ \xi_B\alpha + (1-\xi_B)\beta & \xi_B\beta + (1-\xi_B)\alpha \end{bmatrix}$$

where $\xi_i (i = A, B) = 1$ or 0 represents success or failure of treatment i.

$$H = \begin{bmatrix} \beta p_A + \alpha q_A & \alpha p_A + \beta q_A \\ \alpha p_B + \beta q_B & \beta p_B + \alpha q_B \end{bmatrix}$$

where H is a constant matrix and the maximal eigenvalue is simply the row sum $\lambda = \alpha + \beta$.

Thus, from (1.1) and (1.2), we can obtain results on the proportion of patients allocated to treatment A as

$$\frac{N_{nA}}{n} \xrightarrow{a.s.} v_A = \frac{\alpha p_B + \beta q_B}{\alpha(p_A + p_B) + \beta(q_A + q_B)} \tag{1.4}$$

and the urn composition of type A balls as

$$\frac{Y_{nA}}{Y_{nA} + Y_{nB}} \xrightarrow{a.s.} v_A = \frac{\alpha p_B + \beta q_B}{\alpha(p_A + p_B) + \beta(q_A + q_B)} \tag{1.5}$$

We note that (1.4) is increasing in β/α and tends to $\frac{q_B}{q_A + q_B}$ as $\beta/\alpha \to \infty$. Therefore, if β is large with respect to α, we force the trial to put more patients on the better treatment. But if β/α is too large, the RPW(u, α, β) becomes rather deterministic and may allow unwanted bias in the trial.

1.2.4 Generalizations of GFU Model

Several generalizations of GFU model have been made in recent years. Among them, the first principal one involves allowing the ball selected not to be replaced or allowing some balls to be removed from the urn. Smythe (1996) defined the Extended Pólya Urn model (EPU) with

$$\sum_{q=1}^{K} E(D_{k,\,q}) = c > 0, k = 1, ..., K \tag{1.6}$$

namely, adding an expected constant total number of balls at each stage, but the type k ball drawn does not have to be replaced, and in fact, additional type k balls can be removed from the urn, subjected to (1.6) and a restriction is that one cannot remove more balls of a certain type than are present in the urn (i.e., H is *tenable*).

For the EPU, Smythe (1996) established the asymptotic normality of Y_n and N_n under the assumptions: (i) for each nonprincipal eigenvalues $\lambda_j, \lambda > 2\ \mathrm{Re}(\lambda_j)$; (ii) all eigenvalues are simple, and no two distinct complex eigenvalues have the same real part, except for conjugate pairs; and (iii) the eigenvectors are linearly independent, where $N_n = (N_{n1}, ..., N_{nK})$ and N_{nk} is the number of times a ball of type k drawn in the first n trials.

The second major generalization of the GFU model is the introduction of a nonhomogeneous generating matrix, H_n, where the expected number of balls added to the urn changes across draws. This is the model investigated by Bai and Hu (1999), they assume there is a positive regular matrix H such that

$$\sum_{n=1}^{\infty} n^{-1} ||H_n - H||_{\infty} < \infty \tag{1.7}$$

In this case, the limiting results given in (1.1) and (1.2) remain hold.

From the above introduction of adaptive designs, the GFU model has been playing a significant role in that it can skew the probabilities to favor the treatment that has been the most effective thus far in the trial, thus making the randomization strategy more attractive to physicians than traditional allocations.

We are interested in designs that provide information about dose that maximizes the probability of success, i.e. the *optimal dose*, while treating very few subjects at dosages that have high risks of failure. The aim of this thesis research is to find an optimal dose level for clinical trials through adaptive design using GFU

scheme with consideration of the patients' covariates. In the past literature, the patient's covariate, which actually having effect on the performance of the treatment assigned to the patient, have not been taken into consideration. In Chapter 2, a generalized linear model is proposed with consideration of the patients' covariates and the dose levels simultaneously. The Maximum Likelihood Estimation is used to estimate parameters. The asymptotic properties of the MLE,including the law of large numbers and central limit theorem (CLT) are investigated. A theorem regarding Urn composition is proved. In Chapter 3, a series of simulation is conducted to verify the above results and to select the optimal dose in the circumstance involving patients' covariates. Some discussions and conclusions are given in Chapter 4.

Chapter 2

The New Model

2.1 Introduction to The New Model

Previously, the adaptive designs have not considered the patients' covarients. In those adaptive designs, the performance of treatments is equal for all patients. However, in fact, the effectiveness of treatments should strongly relate to the patients' covariates such as disease history, physical fitness, which will have great effects on the performance of treatments. Here, we are going to take patients' covariates and treatments (or dosage, dose level) into account simultaneously and propose a generalized linear model based on GFU, which can assign more patients to the better treatment only for their specific covariates.

Our model can be described as follows: suppose there are K dose levels denoted

as $d_1, d_2, ..., d_K$. Let X_i for $i = 1, ..., n$ be the i-th patient's covariate, Z_i be the dose level randomly chosen from the K dose levels with certain probabilities. Also, for $k = 1, ..., K$, let $I_{ki} = 1$ if $Z_i = d_k$, and $I_{ki} = 0$ otherwise. Assume that patient's response is dichotomous. Let $Y_i = 1$ if the i-th patient's response is a success, 0 otherwise. Define $p_i = Pr(Y_i = 1|Z_i, X_i)$ for $i = 1, ..., n$ be the probability of success of the i-th patient. Consider the following logistic model:

$$\texttt{logit}p_i = \alpha + \beta Z_i + \gamma Z_i^2 + \lambda X_i \qquad i = 1, ..., n \tag{2.1}$$

Define the allocation function: $F_{k,\ i}(x), k = 1, ..., K$ for the i-th stage. We assume that

$$F_{k,\ 0} \equiv 1, \quad k = 1, ..., K$$

We also assume that the response can be obtained before the next patient enters the trials, i.e. there is no possibility of delayed response. After the response of the i-th patient is obtained, we update the allocation function according to the following equations:

$$F_{k,\ i}(x) = F_{k,\ i-1}(x) + [Y_i I_{k,\ i} + (K-1)^{-1}(1-Y_i)(1-I_{k,\ i})]f(x, x_i) \tag{2.2}$$

where $f(a, b)$ is a decreasing function of $|a - b|$ (a,b are symmetric in f). We also assume that $E_{a,\ b}[(f(a, b))^2] < \infty$. The updating rule implies that if the response of the $(i-1)$-th patient is successful, then we tend to allocate the i-th patient to the same treatment if one has the covariate value near x_{i-1}. Here adding a function

$f(x, x_i)$ to the previous function $F_{k,\ i-1}(x)$ can increase the weight of possibility of allocating the i-th patient with the covariate near to the last covariate x_{i-1} to the more successful dosage.

By(2.2),we have

$$Pr(I_{k,\ i} = 1) = \frac{F_{k,\ i-1}(x_i)}{\sum_{l=1}^{K} F_{l,\ i-1}(x_i)}$$

where $I_{k,\ i} = 1$ if the i-th patient was allocated to dose k, 0 otherwise.

This new model can be seen as an urn model. Suppose an urn with K types of balls. At the i-th stage, the urn composition $(F_{1,\ i}(x), ..., F_{K,\ i}(x))$ is a vector of functions of x. When the new patient enters the trial, then one's covariate x_{i+1} is measured, we plug x_{i+1} into the functions, thus we can get a vector with fixed values: $(F_{1,\ i}(x_{i+1}), ..., F_{K,\ i}(x_{i+1}))$ for $i = 1, ..., n$. If the i-th response is successful on treatment k, then another $f(x, x_i)$ number of type k balls will be added to the urn, otherwise, $(K-1)^{-1} f(x, x_i)$ number of balls will be added to every other type.

2.2 The Likelihood and Asymptotic Properties of MLE

The Maximum Likelihood Estimation (MLE) can be used to estimate the parameters in (2.1). Let $Y^n = \{Y_1, ..., Y_n\}$ be the response history, $Z^n = \{Z_1, ..., Z_n\}$ be the history of design point (treatment) assignment, $X^n = \{X_1, ..., X_n\}$ be the

history of subjects'(patient) covariates.

The likelihood can be written as:

$$\mathscr{L}_n(\theta) = \mathscr{L}_n\{Y^n, Z^n, X^n, \theta\}$$

We assume that:

(1)the response depends on the selected design point, the subjects' covariates, and some parameter vector θ (suppose θ is a vector of dimension s);

(2)future design points are selected according to some function of the data from the response history, design point assignment history and subjects' covariates, but independent of θ;

(3)subjects' covariates are independent.

Consequently, the likelihood can be expressed as follows:

$$\begin{aligned}
\mathscr{L}_n(\theta) &= \mathscr{L}_n\{Y^n, Z^n, X^n; \theta\} \\
&= \mathscr{L}\{Y_n|Y^{n-1}, Z^n, X^n; \theta\}\mathcal{L}\{Z_n|Y^{n-1}, Z^{n-1}, X^n; \theta\} \\
&\quad \mathscr{L}\{X_n|Y^{n-1}, Z^{n-1}, X^{n-1}; \theta\}\mathscr{L}_{n-1}(\theta) \\
&= \mathscr{L}\{Y_n|Z_n, X_n; \theta\}\mathscr{L}\{Z_n|Y^{n-1}, Z^{n-1}, X^n\}\mathscr{L}\{X_n\}\mathscr{L}_{n-1}(\theta)
\end{aligned}$$

The term $\mathscr{L}\{Z_n|Y^{n-1}, Z^{n-1}, X^n\}$ is just the allocation function and is ancillary to the likelihood as is the probability distribution of the covariates. Unwinding the

recursion we can obtain:

$$\begin{aligned}
\mathscr{L}_n(\theta) &\propto \prod_{i=1}^{n} \mathscr{L}\{Y_i|Z_i, X_i;\theta\} \\
&= \prod_{i=1}^{n}\prod_{k=1}^{K} \left[E_{i-1}(I_{k,\ i})\right]^{I_{k,\ i}} p_i^{Y_i I_{k,\ i}} q_i^{(1-Y_i)I_{k,\ i}} \\
&= \prod_{i=1}^{n} \Big(\prod_{k=1}^{K} \left[E_{i-1}(I_{k,\ i})\right]^{I_{k,\ i}}\Big) p_i^{\sum_k Y_i I_{k,\ i}} q_i^{\sum_k (1-Y_i)I_{k,\ i}} \\
&\propto \prod_{i=1}^{n} p_i^{\sum_k Y_i I_{k,\ i}} q_i^{\sum_k (1-Y_i)I_{k,\ i}} \\
&= \prod_{i=1}^{n} p_i^{Y_i} q_i^{1-Y_i} \qquad (2.3)
\end{aligned}$$

By equating the derivative of the log-likelihood with respect to θ to equal 0, we can obtain the maximum likelihood estimator for θ, we denote this estimator as $\hat{\theta}_n$.

It should be noted that $Y_1, ...Y_n$ are dependent random variables due to the sequential design. Consequently, it is necessary to use martingale theory to prove the usual asymptotic properties of maximum likelihood estimators. $\mathscr{L}_n(\theta) \equiv \mathscr{L}_n\{y_1, ..., y_n;\theta\}$ is the joint density of $Y_1, ..., Y_n$, and let

$$\frac{\partial \mathtt{log}\mathscr{L}_n(\theta)}{\partial \theta_a} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_a}\{\mathtt{log}\mathscr{L}_i(\theta) - \mathtt{log}\mathscr{L}_{i-1}(\theta)\} \equiv \sum_{i=1}^{n} \frac{\partial}{\partial \theta_a} U_i(\theta)$$

where $\mathscr{L}_0 = 1, a = 1, ..., s$.

In Appendix 1, a theorem of the limiting results of MLE is proved for the general stochastic process by imposing a certain regularity conditions.

Consider the generalized linear model in the exponential family:

$$f(y_i) = \exp\left\{\frac{y_i\eta_i - a(\eta_i) + b(y_i)}{\phi}\right\}$$

where $\eta_i = h(w_i^T\theta)$ (Nelder and Wedderburn,1972). With this model,

$$E(y_i|w_i) = a^{'}(\eta_i);\ Var(y_i|w_i) = a^{''}(\eta_i)\phi.$$

For the moment, assume that the scale parameter ϕ is fixed. Liking to the notation used in the previous sections, $\theta = (\alpha\ \beta\ \gamma\ \lambda)^T,\ w_i = (1\ z_i\ z_i^2\ x_i)^T$ are 4×1 vectors. And we have:

$$\frac{\partial U_i(\theta)}{\partial\theta_a} = \frac{1}{\phi}\{y_i - a^{'}(\eta_i)\}h^{'}(w_i^T\theta)w_{ai} \tag{2.4}$$

Theorem 1 *Under the following conditions:*

$$\frac{1}{n\phi}\sum_{i=1}^{n} E_{i-1}\Big([h^{'}(w_i^T\theta)]^2 w_{ai}w_{bi}a^{''}(\eta_i)\Big) \xrightarrow{P} \gamma_{ab}\ (as\ n\to\infty), a, b = 1, ..., 4, \tag{2.5}$$

where $\Gamma = (\gamma_{ab})$ *is nonsingular,*

$$\frac{1}{n^{3/2}\phi}\sum_{i=1}^{n} E_{i-1}\Big([h^{'}(w_i^T\theta)]^3 w_{ai}^3 a^{'''}(\eta_i)\Big) \xrightarrow{P} 0 \quad (as\ n\to\infty), \tag{2.6}$$

$$\frac{1}{n^2\phi}\sum_{i=1}^{n} [h^{''}(w_i^T\theta)]^2 w_{ai}^2 w_{bi}^2 a''(\eta_i) \xrightarrow{P} 0 \qquad (as\ n\to\infty), \tag{2.7}$$

the solution $\hat{\theta}_n$ *to the score function*

$$\sum_{i=1}^{n}\frac{\partial U_i(\theta)}{\partial\theta_a} = 0 \tag{2.8}$$

is consistent for θ, *and* $n^{1/2}(\hat{\theta} - \theta) \to N(0, \Gamma^{-1})$.

The theorem results directly from conditions(A1)-(A5) of Appendix 1. (A1) and (A2) are standard regularity conditions that apply to exponential families. Condition (2.5) derives from (A3).

Note that

$$\begin{aligned} & E_{i-1}\Big(\{h'(w_i^T\theta)\}^2 w_{ai} w_{bi}\{y_i - a'(\eta_i)\}^2 | x_i, z_i\Big) \\ = \; & E_{i-1} E\{h'(w_i^T\theta)\}^2 w_{ai} w_{bi} E\Big(\{y_i - a'(\eta_i)\}^2 | x_i, z_i\Big) \\ = \; & E_{i-1}\Big(\{h'(w_i^T\theta)\}^2 w_{ai} w_{bi} a''(\eta_i)\Big) \end{aligned}$$

Then the left-hand side of (A3) can be written as

$$n^{-1}\sum_{i=1}^{n} \frac{1}{\phi^2} E_{i-1}\Big(\{h'(w_i^T\theta)\}^2 w_{ai} w_{bi}\{y_i - a'(\eta_i)\}^2\Big)$$

which then can be written as the left-hand side of (2.5). Condition (2.6) implies condition (A4) with $\delta = 1$, and conditions (2.5) and (2.7), together with (A4), implies (A6).

Let $h(t) = t, a(\eta) = \texttt{log}(1 + e^{\eta})$, and $\phi = 1$. Then $h^{'} = 1, h'' = 0$, and (2.7) is automatically satisfied. Condition (2.6) reduces to show that

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n} E_{i-1}\Big(w_{ai}^3 \frac{e^{\eta_i}(1 - e^{\eta_i})}{(1 + e^{\eta_i})^3}\Big) \xrightarrow{P} 0 \quad (as\ n \to \infty)$$

and this follows since the summands are bounded in i. Condition (2.5) establishes the variance-covariance structure via

$$\frac{1}{n}\sum_{i=1}^{n} E_{i-1}\Big(w_{ai} w_{bi} p_i(1 - p_i)\Big) \xrightarrow{P} \gamma_{ab} \quad (as\ n \to \infty) \tag{2.9}$$

2.3 Asymptotic properties of Urn composition

The asymptotic property of Urn composition is a main concern when investigating statistically the Urn model.

From the model stated in Section 2.1, we have:

$$\sum_{k=1}^{K} F_{k,\ n}(x) = K + \sum_{i=1}^{n} f(x, x_i)$$

Let $g(x)$ is the expectation of $f(x, x_i)$ with respect to X_i. Because $X_1, ..., X_n$ are i.i.d. random variables, from the Law of Large Numbers, we have:

$$\frac{\sum_{i=1}^{n} f(x, x_i)}{n} \overset{a.s.}{\to} g(x) \quad (as\ n \to \infty)$$

then, as $n \to \infty$:

$$\frac{\sum_{k=1}^{K} F_{k,\ n}(x)}{n} = K/n + n^{-1} \sum_{i=1}^{n} f(x, x_i) \overset{a.s.}{\to} g(x) \tag{2.10}$$

and

$$E_x \left[\frac{\sum_{k=1}^{K} F_{k,\ n}(x)}{n} \right] \to E_x g(x) \equiv g \tag{2.11}$$

Denote $A_{k,\ i} \equiv [Y_i I_{k,\ i} + (K-1)^{-1}(1-Y_i)(1-I_{k,\ i})] f(x, x_i)$ and let $\mathscr{F}_n \equiv \sigma\{Y^n, X^{n+1}, Z^n\}$ be the $\sigma-$field after n stages.Then,

$$\begin{aligned} F_{k,\ n}(x) &= 1 + \sum_{i=1}^{n} [Y_i I_{k,\ i} + (K-1)^{-1}(1-Y_i)(1-I_{k,\ i})] f(x, x_i) \\ &= 1 + \sum_{i=1}^{n} A_{k,\ i} \\ &= 1 + \sum_{i=1}^{n} E\big(A_{k,\ i} \big| \mathscr{F}_{i-1}\big) + \sum_{i=1}^{n} \big[A_{k,\ i} - E\big(A_{k,\ i} \big| \mathscr{F}_{i-1}\big)\big] \end{aligned}$$

where $\{A_{k,\ i} - E(A_{k,\ i}|\mathscr{F}_{i-1}), \mathscr{F}_i, i = 1, ..., n\}$ is a martingale difference because:

$$E[A_{k,\ i} - E(A_{k,\ i}|\mathscr{F}_{i-1})|\mathscr{F}_{i-1}] = 0$$

Let $p_{ki} = 1 - q_{ki}$ be the probability of success if the i-th patient is allocated to the k-th treatment.

$$\begin{aligned}
\sum_{i=1}^{n} E[A_{k,\ i}|\mathscr{F}_{i-1}] &= \sum_{i} f(x, x_i)\left[\frac{p_{ki}F_{k,\ i-1}(x_i)}{\sum_{l=1}^{K} F_{l,\ i-1}(x_i)} + (K-1)^{-1}\frac{\sum_{l\neq k} q_{li}F_{l,\ i-1(x_i)}}{\sum_{l=1}^{K} F_{l,\ i-1}(x_i)}\right] \\
&= \sum_{i=1}^{n} f(x, x_i)\frac{[(K-1)p_{ki} - q_{ki}]F_{k,\ i-1}(x_i) + \sum_{l=1}^{K} q_{li}F_{l,\imath-1}(x_i)}{(K-1)[K + \sum_{j}^{i-1} f(x_i, x_j)]} \\
&= f(x, x_1)\frac{Kp_{k1} + \sum_{l=1}^{K} q_{l1} - 1}{K(K-1)} \\
&+ \sum_{i=1}^{n} f(x, x_i)\frac{[Kp_{ki} - 1]F_{k,\ i-1}(x_i) + \sum_{l=1}^{K} q_{li}F_{l,\ i-1}(x_i)}{(K-1)[K + \sum_{j}^{i-1} f(x_i, x_j)]} \\
&= \delta_0 + \sum_{i=2}^{n} f(x, x_i)\frac{[Kp_{ki} - 1]F_{k,\ i-1}(x_i) + \sum_{l} q_{li}F_{l,\ i-1}(x_i)}{(K-1)[K + (i-1)g(x_i)]} + \delta_{1n}
\end{aligned}$$

where

$$\delta_0 \equiv f(x, x_i)\frac{Kp_{k1} + \sum_{l=1}^{K} q_{l1} - 1}{K(K-1)}$$

and

$$\begin{aligned}
\delta_{1n} \equiv \sum_{i=1}^{n}(K-1)^{-1} f(x, x_i)\Big\{[Kp_{ki} - 1]F_{k,\ i-1}(x_i) + \sum_{l} q_{li}F_{l,\ i-1}(x_i)\Big\} \\
\cdot\left\{\frac{1}{K + \sum_{j=1}^{i-1} f(x_i, x_j)} - \frac{1}{K + (i-1)g(x_i)}\right\}
\end{aligned}$$

note that both δ_0, and δ_{1n} are related to k.

Take expectation of $\sum_{i=1}^{n} E[A_{k,\ i}|\mathscr{F}_{i-1}]$ with respect to X, here, we denote $\int_x \cdot$

as the expectation with respect to X. we obtain:

$$
\begin{aligned}
& \int_x \sum_{i=1}^{n} E\big[A_{k,\ i}\big|\mathscr{F}_{i-1}\big] \\
= & \int_x \delta_0 + \int_x \delta_{1n} + \int_x \sum_{i=2}^{n} f(x,x_i)\frac{(Kp_{ki}-1)F_{k,\ i-1}(x_i)+\sum_l q_{li}F_{l,\ i-1}(x_i)}{(K-1)[K+(i-1)g(x_i)]} \\
= & \ g(x_1)\frac{Kp_{k1}+\sum_l q_{li}-1}{K(K-1)} + \sum_{i=2}^{n} g(x_i)\frac{(Kp_{ki}-1)F_{k,\ i-1}(x_i)+\sum_l q_{li}F_{l,\ i-1}(x_i)}{(K-1)[K+(i-1)g(x_i)]} \\
+ & \ \sum_{i=1}^{n}\frac{g(x_i)}{K-1}\Big\{[Kp_{ki}-1]F_{k,\ i-1}(x_i)+\sum_l q_{li}F_{l,\ i-1}(x_i)\Big\} \\
& \qquad\qquad \cdot\Big\{\frac{1}{K+\sum_{j=1}^{i-1}f(x_i,x_j)} - \frac{1}{K+(i-1)g(x_i)}\Big\} \\
\equiv & \ \Delta_0+\Delta_n+\Delta_{1n}
\end{aligned}
$$

where $\Delta_0, \Delta_n, \Delta_{1n}$ denote the above three corresponding terms.

Then,

$$
\begin{aligned}
\int_x F_{k,\ n}(x) &= 1+\int_x \sum_{i=1}^{n} E\big(A_{k,\ i}\big|\mathscr{F}_{i-1}\big) + \int_x \sum_{i=1}^{n}\big[A_{k,\ i}-E\big(A_{k,\ i}\big|\mathscr{F}_{i-1}\big)\big] \\
&= 1+\Delta_0+\Delta_n+\Delta_{1n}+\Delta_{2n}
\end{aligned}
$$

where

$$
\Delta_{2n} \equiv \int_x \sum_{i=1}^{n}\big[A_{k,\ i}-E\big(A_{k,\ i}\big|\mathscr{F}_{i-1}\big)\big]
$$

The term

$$
\Delta_n = \sum_{i=1}^{n}\frac{g(x_i)}{(K-1)[K+(i-1)g(x_i)]}[(Kp_{ki}-1)F_{k,\ i-1}(x_i)+\sum_l q_{li}F_{l,\ i-1}(x_i)]
$$

$$
\begin{aligned}
&= \sum_{i=2}^{n} \frac{1}{(K-1)(i-1)}[(Kp_{ki}-1)F_{k,\ i-1}(x_i) + \sum_l q_{li}F_{k,\ i-1}(x_i)] \\
&\quad - \sum_{i=1}^{n} \frac{K}{(K-1)(i-1)[K+(i-1)g(x_i)]}[(Kp_{ki}-1)F_{k,\ i-1}(x_i) + \sum_l q_{li}F_{l,\ i-1}(x_i)] \\
&\equiv R_{1n} - R_{2n}
\end{aligned}
$$

where R_{1n}, R_{2n} denote the above two corresponding terms.

Let $R_n \equiv \Delta_{1n} + \Delta_{2n} - R_{2n}$, then:

$$
\begin{aligned}
\int_x F_{k,\ n}(x) &= 1 + \Delta_0 + R_{1n} - R_{2n} + \Delta_{1n} + \Delta_{2n} \\
&= 1 + \Delta_0 + R_{1n} + R_n \\
&= 1 + \Delta_0 + ER_{1n} + (R_{1n} - ER_{1n}) + R_n
\end{aligned}
$$

where the expectation is with respect to X_i.

Suppose $\texttt{Cov}[f(x_i, x_j), p_i \big| \mathscr{F}_{i-1}] = 0$ for $k = 1, ...K, j < i$, then

$$
\begin{aligned}
&\texttt{Cov}[F_{k,\ i-1}(x_i), Kp_i - 1] \\
&= E\Big[\texttt{Cov}(F_{k,\ i-1}(x_i), Kp_i - 1)\Big|\mathscr{F}_{i-1}\Big] \\
&= E\Big[\sum_{j=1}^{i-1} K\Big(Y_j I_{k,\ j} + (K-1)^{-1}(1-Y_j)(1-I_{k,\ j})\Big)\texttt{Cov}[f(x_i, x_j), p_i\big|\mathscr{F}_{i-1}]\Big] \\
&= 0
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& ER_{1n} = \int_{x_i} R_{1n} \\
= & \sum_{i=2}^{n} \frac{1}{(K-1)(i-1)} \Big[\int_{x_i} F_{k,\ i-1}(x_i) \cdot \int_{x_i} (Kp_{ki} - 1) + \sum_{l} \int_{x_i} F_{l,\ i-1}(x_i) \cdot \int_{x_i} q_{li} \Big] \\
& + \sum_{i=2}^{n} \frac{1}{(K-1)(i-1)} \Big\{ \mathtt{Cov}[F_{k,\ i-1}(x_i), Kp_{ki} - 1] + \sum_{l} \mathtt{Cov}[F_{l,\ i-1}(x_i), q_{li}] \Big\} \\
= & \sum_{i=2}^{n} \frac{1}{(K-1)(i-1)} \Big[\int_{x_i} F_{k,\ i-1}(x_i) \cdot \int_{x_i} (Kp_{ki} - 1) + \sum_{l} \int_{x_i} F_{l,\ i-1}(x_i) \cdot \int_{x_i} q_{li} \Big]
\end{aligned}
$$

Denote $W_n \equiv (R_{1n} - ER_{1n}) + R_n$, then,

$$
\begin{aligned}
& \int_{x} F_{k,\ n} \\
= & \ 1 + \Delta_0 + W_n \\
& + \sum_{i=2}^{n} \frac{1}{(K-1)(i-1)} \left[\int_{x_i} F_{k,\ i-1}(x_i) \cdot \int_{x_i} (Kp_{ki} - 1) + \sum_{l} \int_{x_i} F_{l,\ i-1}(x_i) \cdot \int_{x_i} q_{li} \right]
\end{aligned}
$$

Denote $\Delta W_n \equiv W_n - W_{n-1}$, therefore

$$
\begin{aligned}
& \int_{x} F_{k,\ n} - \int_{x} F_{k,\ n-1} \\
= & \ \frac{1}{(K-1)(n-1)} \Big[\int_{x_n} F_{k,\ n-1}(x_n) \cdot \int_{x_n} (Kp_{kn} - 1) + \sum_{l=1}^{K} \int_{x_n} F_{l,\ n-1}(x_n) \cdot \int_{x_n} q_{ln} \Big] + \Delta W_n
\end{aligned}
$$

Let $P_k \equiv \int_{x_i} p_{ki} = 1 - Q_k$, Write the above as a vector, we obtain:

$$\begin{bmatrix} \int_x F_{1,\ n} - \int_x F_{1,\ n-1} \\ \int_x F_{2,\ n} - \int_x F_{2,\ n-1} \\ \cdots \\ \cdots \\ \cdots \\ \int_x F_{K,\ n} - \int_x F_{K,\ n-1} \end{bmatrix}$$

$$= \frac{1}{(K-1)(n-1)} \begin{pmatrix} (K-1)P_1 & Q_2 & \cdots & Q_K \\ Q_1 & (K-1)P_2 & \cdots & Q_K \\ \cdots & \cdots & \cdots & \cdots \\ Q_1 & Q_2 & \cdots & (K-1)P_K \end{pmatrix} \cdot \begin{bmatrix} \int_{x_n} F_{1,\ n-1}(x_n) \\ \int_{x_n} F_{2,\ n-1}(x_n) \\ \cdots \\ \cdots \\ \cdots \\ \int_{x_n} F_{K,\ n-1}(x_n) \end{bmatrix}$$

$$+ \quad \mathbf{\Delta W_n}$$

Denote

$$\mathbf{H} \equiv \begin{pmatrix} (K-1)P_1 & Q_2 & \cdots & Q_K \\ Q_1 & (K-1)P_2 & \cdots & Q_K \\ \cdots & \cdots & \cdots & \cdots \\ Q_1 & Q_2 & \cdots & (K-1)P_K \end{pmatrix} \Big/ (K-1) \tag{2.12}$$

and

$$\overline{\mathbf{F}}_n \equiv \begin{bmatrix} \int_x F_{1,\,n} \\ \int_x F_{2,\,n} \\ \cdots \\ \cdots \\ \cdots \\ \int_x F_{K,\,n} \end{bmatrix}$$

then we have

$$\overline{\mathbf{F}}_n = (\mathbf{I} + \frac{1}{n-1}\mathbf{H})\overline{\mathbf{F}}_{n-1} + \mathbf{\Delta W_n} \tag{2.13}$$

Assumption 1 *Suppose that* $\mathbf{H}$ *is of non-negative entries and* $\mathbf{H}^{'}$ *(transpose of* $\mathbf{H}$*)has the Jordan form decomposition*

$$\mathbf{T^{-1}H^{'}T} = \mathbf{J} = \begin{pmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \mathbf{J_1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{J_r} \end{pmatrix} \quad \textbf{with } \mathbf{J_t} = \begin{pmatrix} \lambda_t & 1 & 0 & \cdots & 0 \\ 0 & \lambda_t & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_t & 1 \\ 0 & 0 & 0 & \cdots & \lambda_t \end{pmatrix}$$

where $\lambda_0 = 1$ *is the unique maximum eigenvalue of* $\mathbf{H}$*. Moreover, we assume that the elements of the eigenvector* $\mathbf{v} = (v_1, ..., v_K)^T$ *associated with the positive maximal eigenvalue* λ_0 *are nonnegative and satisfy* $\sum_{k=1}^{K} v_k = 1$*.*

Theorem 2 *As* $n \to \infty$*, we have the following asymptotic property:*

$$F_{k,\ n}(x)/n \xrightarrow{a.s.} g(x)v_k$$

And the Urn Composition:

$$\frac{F_{k,\ n}(x)}{\sum_{l=1}^{K} F_{l,\ n}(x)} \xrightarrow{a.s.} v_k \tag{2.14}$$

Moreover, the proportion of patients allocated to a dose level has the following result:

$$n^{-1} \sum_{i=1}^{n} I_{k,\ i} \xrightarrow{a.s.} v_k \tag{2.15}$$

for $k = 1, ..., K$*.*

See Appendix 2 for the Proof.

These results accord with those in the usual GFU model.

Chapter 3

Simulation Results

To verify the asymptotic results obtained is Chapter 2 and help with selection of optimal dose, a series of simulations are conducted. R language is used to conduct these simulations. A typical source code can be found in Appendix 3.

3.1 Algorithm

The algorithm can be described as the following procedures:

1 Set K=10 dose levels, they are $d_k = k/10, k = 1, ..., 10$. Generate a sequence of random variables $X_i, i = 1, ..., n$ from a standard normal distribution or uniform distribution over (0,1), then set some values for the parameters in (2.1).

2 For the first trial, we allocate the patient to a certain dose level with equal probability, i.e. $Pr(Z_1 = d_k) = 0.1, k = 1, ..., K$. In the following trials, for the i-th patient, we plug the patient's covariate X_i into the allocation function, then generate the allocation function $I_{k,\ i}, k = 1, ..., K$ with probabilities $Pr(I_{k,\ i} = 1) = \frac{F_{k,\ i-1}(x_i)}{\sum_{j=1}^{K} F_{j,\ i-1}(x_i)}, k = 1, ..., K$. The allocation function $F_{k,\ i}(x), k = 1, ..., K$ is given by (2.2).

3 By the logistic regression model, we can obtain the response of each patient:$Y_i(i = 1, ..., n) = 1$ or 0 represents success or failure of patient i. Y_i is generated with probabilities $Pr(Y_i = 1) = p_{k,\ i}$ if dose k is assigned to patient i.

4 According to (2.3), we can obtained MLE for the parameters the same as in i.i.d case using the generated n responses. A circle is finished till now.

5 In order to investigate the asymptotic properties of MLE, we conduct N circles to obtain N parameter estimations.

6 To investigate the asymptotic properties of Urn composition and select of optimal dose, one circle with very large n is chosen.

3.2 Simulation Results

3.2.1 Asymptotics of Maximum Likelihood Estimation (MLE)

Our investigations of normality will concentrate on the behavior of observations in one or two dimensions. As might be expected, it can be proved difficult to construct a "good" overall test of joint normality in more than two dimensions because many aspects may be wrong. To some extent, we must pay a price for concentrating on univariate and/or bivariate examinations of normality: We can never be sure that we have not missed some feature that is revealed only in higher dimensions. For convenience, we only check one dimension asymptotic behavior of MLE. In R language, Shapiro-Wilk (1965) test is provided to test one dimension normality. In Table 3.1-Table 3.4, we set $\alpha = -0.5, \beta = 10, \gamma = -10, \lambda = 2$ (in 2.1), and let $f(a,b) = \dfrac{1}{1+|a-b|}$ or $f(a,b) = \dfrac{1}{1+|a-b|^2}$ and X_i's be from Standard Normal Distribution or Uniform Distribution over (0,1). Let $n = 500$ and $N = 100$. In these 4 cases, the performances of MLEs are good.The tests are not significant as n is large(>200). A more detailed simulation results are presented in Table 3.5 and Table 3.6.

Table 3.1: $f(a,b) = \dfrac{1}{1+|a-b|}, n = 500, N = 100, X \sim N(0,1)$.

Parameter	True Value	Mean	Variance	p-value
α	-0.5	-0.49	0.301	0.18
β	10	10.16	4.537	0.37
γ	-10	-10.26	3.597	0.64
λ	2	2.03	0.045	0.57

Table 3.2: $f(a,b) = \dfrac{1}{1+|a-b|^2}, n = 500, N = 100, X \sim N(0,1)$.

Parameter	True Value	Mean	Variance	p-value
α	-0.5	-0.55	0.238	0.31
β	10	10.18	3.865	0.42
γ	-10	-10.15	3.049	0.53
λ	2	1.98	3.044	0.77

Table 3.3: $f(a,b) = \dfrac{1}{1+|a-b|}, n = 500, N = 100, X \sim U(0,1)$.

Parameter	True Value	Mean	Variance	p-value
α	-0.5	-0.62	0.368	0.67
β	10	10.24	5.806	0.24
γ	-10	-10.20	4.549	0.20
λ	2	2.08	0.283	0.68

Table 3.4: $f(a,b) = \dfrac{1}{1+|a-b|^2}, n = 500, N = 100, X \sim U(0,1)$.

Parameter	True Value	Mean	Variance	p-value
α	-0.5	-0.58	0.363	0.70
β	10	10.33	6.040	0.80
γ	-10	-10.28	4.750	0.46
λ	2	2.06	0.282	0.47

Table 3.5: Comparisons on Different Situations-1,$X \sim N(0,1)$.

Parameter	TrueValue	Mean	VAR				p-value
Case 1-1	n=200	N=100					
α	-0.5	-0.44	0.764	-2.950	2.457	-0.027	0.36
β	10	10.13	2.950	14.195	12.815	0.407	0.42
γ	-10	-10.24	2.457	-12.815	12.092	-0.395	0.23
λ	2	2.08	-0.027	0.407	-0.395	0.111	0.78
Case 1-2	n=200	N=400					
α	-0.5	-0.54	0.557	-2.208	1.801	-0.011	0.75
β	10	10.35	-2.208	11.198	-10.020	0.316	0.22
γ	-10	-10.39	1.801	-10.020	9.416	-0.334	0.12
λ	2	2.082	-0.011	0.316	-0.334	0.107	0.04
Case 1-3	n=500	N=100					
α	-0.5	-0.49	0.304	-1.042	0.783	0.020	0.18
β	10	10.16	-1.042	4.537	-3.888	0.031	0.37
γ	-10	-10.26	0.783	-3.888	3.597	-0.064	0.64
λ	2	2.02	0.020	0.031	-0.064	0.045	0.57
Case 1-4	n=500	N=400					
α	-0.5	-0.48	0.245	-0.945	0.763	0.005	0.43
β	10	10.04	-0.945	4.479	-3.921	0.063	0.68
γ	-10	-10.07	0.763	-3.921	3.594	-0.069	0.77
λ	2	2.04	0.005	0.063	-0.069	0.037	0.12
Case 1-5	n=1000	N=100					
α	-0.5	-0.54	0.127	-0.534	0.458	-0.008	0.75
β	10	10.11	-0.534	2.652	-2.395	0.074	0.27
γ	-10	-10.10	0.458	-2.395	2.238	-0.072	0.34
λ	2	2.01	-0.008	0.074	-0.072	0.021	0.09
Case 1-6	n=1000	N=400					
α	-0.5	-0.51	0.099	-0.396	0.330	0.000	0.78
β	10	10.05	-0.396	2.009	-1.823	0.045	0.19
γ	-10	-10.04	0.330	-1.823	1.730	-0.046	0.11
λ	2	2.01	0.000	0.045	-0.046	0.018	0.14

Table 3.6: Comparisons on Different Situations-2,$X \sim U(0,1)$.

Parameter	TrueValue	Mean	VAR				p-value
Case 2-1	n=200	N=100					
α	1	1.17	0.902	-3.232	2.579	0.052	0.085
β	8	7.82	-3.232	16.500	-14.686	-0.022	0.396
γ	-8	-7.85	2.579	-14.686	13.772	0.033	0.457
λ	1	1.031	0.052	-0.022	0.033	0.105	0.996
Case 2-2	n=200	N=400					
α	1	1.00	0.934	-3.459	2.736	0.029	0.001
β	8	8.436	-3.459	16.475	-14.251	0.093	0.660
γ	-8	-8.449	2.736	-14.251	12.878	-0.100	0.539
λ	1	1.03	0.029	0.093	-0.100	0.076	0.030
Case 2-3	n=500	N=100					
α	1	1.05	0.289	-1.060	0.867	0.029	0.809
β	8	8.00	-1.060	5.425	-4.893	-0.040	0.760
γ	-8	-8.04	0.867	-4.893	4.626	0.038	0.975
λ	1	1.04	0.029	-0.040	0.038	0.031	0.070
Case 2-4	n=500	N=400					
α	1	0.97	0.306	-1.164	0.915	0.001	0.57
β	8	8.39	-1.164	5.735	-4.956	0.092	0.44
γ	-8	-8.39	0.915	-4.956	4.491	-0.089	0.41
λ	1	1.03	0.001	0.092	-0.089	0.029	0.25
Case 2-5	n=1000	N=100					
α	1	1.02	0.143	-0.527	0.409	0.002	0.25
β	8	7.92	-0.527	2.752	-2.357	0.053	0.09
γ	-8	-7.92	0.409	-2.357	2.110	-0.050	0.20
λ	1	1.00	0.002	0.053	-0.050	0.013	0.75
Case 2-6	n=1000	N=400					
α	1	1.01	0.181	-0.675	0.531	0.002	0.83
β	8	7.97	-0.675	3.218	-2.77	0.036	0.23
γ	-8	-7.96	0.531	-2.773	2.501	-0.034	0.30
λ	1	1.01	0.002	0.036	-0.034	0.014	0.50

3.2.2 Asymptotics of Urn Composition and Allocation Proportion

By theorem 2,we know that the Urn composition converge to an eigenvector described in Assumption 1. Table-3.7 gives the result as $\alpha = -0.5, \beta = 10, \gamma = -10, \lambda = 2, f(a,b) = \dfrac{1}{1+|a-b|}, X \sim N(0,1)$:

Table 3.7: Urn composition and allocation proportion Convergence, n=20,000

Dose Level	Eigenvector v_k	$\dfrac{F_{k,\ n}(x)}{\sum_{j=1}^{K} F_{j,\ n}(x)}$	$n^{-1}\sum_{i=1}^{n} I_{k,\ i}$	$g(x)v_k$	$F_{k,\ n}(x)/n$
0.1	0.0700	0.0713	0.074	0.0336	0.0343
0.2	0.0885	0.0956	0.087	0.0425	0.0460
0.3	0.1079	0.0917	0.101	0.0518	0.0441
0.4	0.1309	0.1280	0.123	0.0629	0.0616
0.5	0.1423	0.1345	0.135	0.0684	0.0647
0.6	0.1299	0.1311	0.124	0.0624	0.0630
0.7	0.1117	0.1142	0.114	0.0537	0.0549
0.8	0.0948	0.1081	0.108	0.0455	0.0520
0.9	0.0718	0.0745	0.076	0.0345	0.0358
1.0	0.0520	0.0507	0.055	0.0249	0.0244

From the above table, we can find that, for a random generated value x (for convenience, the last patient's covariate is used), both of the Urn composition and allocation proportion near to $\mathbf{v}$ as n is very large. We can also see that dose level 0.5 is the optimal dose with its highest Urn composition and allocation proportion, thus, make more patients be allocated to the optimal dose.

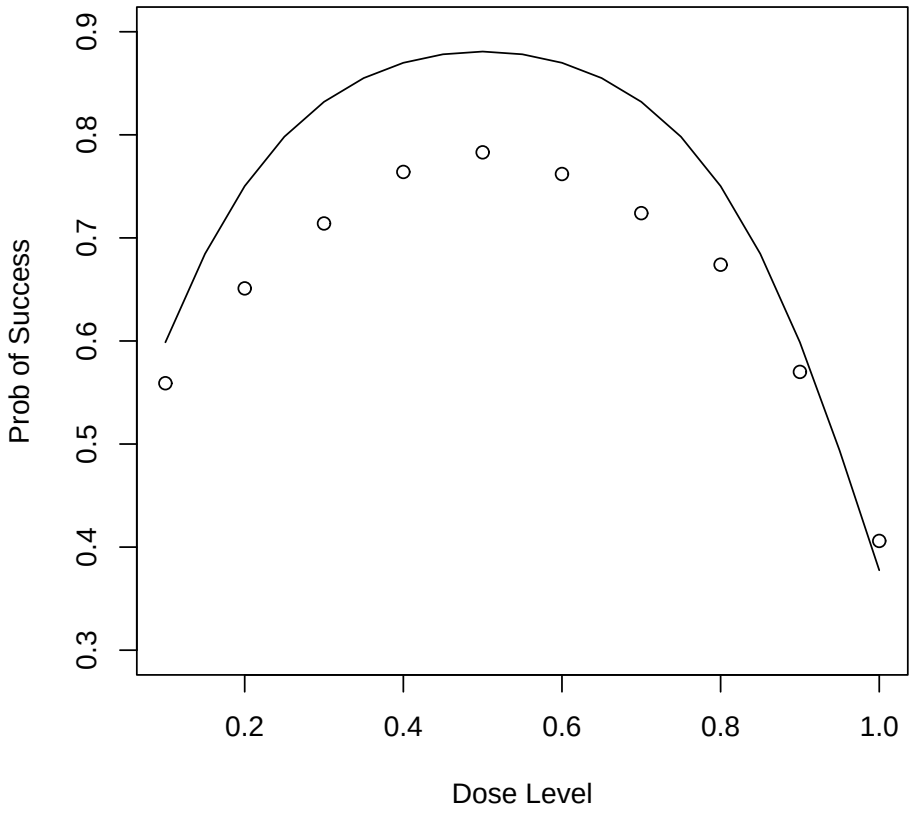
3.2.3 Selection of Optimal Dose

The patients' covariates may have, in certain extent, positive or negative effects on the probability of success of the dose level assigned to this patient. If the coefficient λ in the GLM is relatively great, the effects on the probabilities of success will adverse the selection of optimal dose, because of the randomness of the patients' covariates. Let n be very large(>1,000), this randomness can be counteracted if each dose is drawn enough times. Taking average of the probabilities of success for each dose over the number of patients allocated to, we obtain an approximation of the probability of success for each dose. Table-3.8 is an example in which we use: $\alpha = -0.5, \beta = 10, \gamma = -10, \lambda = 2, f(a,b) = \dfrac{1}{1+|a-b|}, X \sim N(0,1), n = 20,000$. The probabilities of success for each dose when taking patients' covariates into account, still favor the optimal dose we prescribed when not considering the patients' effects. Also, the number of times of allocation has the same trend with the probabilities of success and shows that this model can actually take the patients' effects into account and give an ethical concern. The probabilities of success for each dose when taking patients' covariates into account have, on average, the same trend as without patients' effect. This can be seen more clearly in Figure-3.1, where the solid line is the probabilities of success without effect of patients' covariates and the dots are the probabilities of success with effect of patients' covariates.

Table 3.8: The effect of patients on the dose probability success.

DoseLevel	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Prob.ofSucc.	0.559	0.651	0.714	0.764	0.783	0.762	0.724	0.674	0.570	0.406
No. of Alloc.	1481	1749	2024	2467	2702	2477	2279	2179	1527	1115

Figure 3.1: The Effect of Patients' Covariates on Dose Selection



Chapter 4

Conclusions and Discussions

The main results of this assay are the asymptotic properties of Urn composition and allocation proportion in the circumstance of considering the patients' covariates' effects on the adding rule and the probability of success for each dose. The ethic imperative are given consideration without affecting the statistical inference of the model. As a result, this model places more patients to the optimal dose and lowers the patients's exposition to dose level with high risks of failure. In our model, the generating matrix is in fact $f(x, x_i)H^{'}$. If we think that $f(x, x_i)$ is associated with a weight to increase the probabilities of the allocating the i-th patient with the covariate near to the last covariate x_{i-1} to the more successful dosage, then $H^{'}$ plays a role as a "generating matrix" as is in the GFU model. The asymptotic properties of Urn composition and allocation proportion are related to the unit eigenvector of the maximal eigenvalue of H, as is in the usual GFU case without considering the patients' covariates.

Appendix 1

Let $\mathscr{F}_n$ be the σ-field generated by the stochastic process through stage n ($\mathscr{F}_n \equiv \sigma\{Y^n, X^n, Z^n\}$,in this thesis, denoting $\mathscr{F}_n \equiv \sigma\{Y^n, X^{n+1}, Z^n\}$ is suitable), and define the conditional expectation of $\cdot$ with respect to $\mathscr{F}_n$ to be $E_n(\cdot)$ and the conditional variance to be $Var_n(\cdot)$. We will suppose there exists an open subset Ω_0 of the parameter space Ω containing the true parameter.

We impose the following regularity conditions on the likelihood:

(A1) $\int \mathscr{L}_n(y_1, ..., y_n; \theta) dy_n$ can be partially differentiated twice with respect to θ under the integral sign and the first partials have finite moments of order $2 + \delta$ for some $\delta > 0$.

(A2) For almost all $y_1, ..., y_n$, $\mathscr{L}_n(\theta)$ admits all third partial derivatives, and the absolute values of the third partials (with respect to $\theta_a, \theta_b, \theta_c$) are bounded by a function $M_n(y_1, ..., y_n)$ for all $\theta \in \Omega_0$. We assume $M_{abc} \equiv \sup_n M_n(y_1, ..., y_n)$ is integrable.

(A3) For $a = 1, ..., s, b = 1, ..., s$,

$$n^{-1} \sum_{i=1}^{n} E_{i-1}\{\frac{\partial}{\partial\theta_a} U_i(\theta) \ \frac{\partial}{\partial\theta_b} U_i(\theta)\} \xrightarrow{P} \gamma_{ab}(\theta) \quad (as \ n \to \infty)$$

where $\gamma_{ab}(\theta)$ is a nonrandom function of θ, for all $\theta \in \Omega_0$.

(A4) For some $\delta > 0$ and $a = 1, ..., s$,

$$n^{-(1+\delta/2)} \sum_{i=1}^{n} E_{i-1}\{\frac{\partial}{\partial\theta_a} U_i(\theta)\}^{2+\delta} \xrightarrow{P} 0, \quad (as \ n \to \infty)$$

for all $\theta \in \Omega_0$.

(A5) For $a = 1, ..., s, b = 1, ..., s$,

$$n^{-1} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_a \partial\theta_b} U_i(\theta) \xrightarrow{P} -\gamma_{ab}(\theta) \quad (as \ n \to \infty)$$

for all $\theta \in \Omega_0$, where $\gamma_{ab}(\theta)$ is defined in (A3).

Define $\Gamma(\theta)$ to be an $s \times s$ matrix with elements $\gamma_{ab}(\theta)$, where the γ_{ab}'s are defined in the condition (A3). Let $\hat{\theta}_n \equiv (\hat{\theta}_{1n}, ..., \hat{\theta}_{sn})$ be a MLE for θ. We have the following theorem:

Theorem 3 *If conditions(A1)-(A5) are satisfied, then a consistent MLE $\hat{\theta}_n$ exists and the vector $n^{1/2}(\hat{\theta}_n - \theta)$ is asymptotically multivariate normal with mean zero and variance-covariance matrix $[\Gamma(\theta)]^{-1}$, provided the inverse exists.*

Proof: Let $L_n(\theta) \equiv \texttt{log}\mathscr{L}_n(\theta) = \sum_{i=1}^{n} U_i(\theta)$ be the log-likelihood, Suppose θ_0 is the true parameter,using Taylor's Expansion, we have:

$$0 = L_n'(\hat{\theta}_n) = L_n'(\theta_0) + L_n''(\theta_1)(\hat{\theta}_n - \theta_0)$$

where L_n' is a $s \times 1$ vector and L_n'' is a $s \times s$ matrix, θ_1 is a vector among two balls with radii $||\theta_0||$ and $||\hat{\theta}_n||$.

Then

$$\hat{\theta}_n - \theta_0 = -[L_n''(\theta_1)]^{-1} L_n'(\theta_0) = -\Big[\frac{L_n''(\theta_1)}{n}\Big]^{-1} \frac{L_n'(\theta_0)}{n} \tag{4.1}$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\Big[\frac{L_n''(\theta_1)}{n}\Big]^{-1} \frac{L_n'(\theta_0)}{\sqrt{n}} \tag{4.2}$$

From (A1), we have

$$\begin{aligned} E_{i-1}\big[U_i'(\theta)\big] &= \int \frac{\partial f(y_i|\mathscr{F}_{i-1},\theta)}{\partial\theta} dy_i = \int \Big(\frac{\mathscr{L}_i}{\mathscr{L}_{i-1}}\Big)' dy_i \\ &= \int \frac{\mathscr{L}_i' \mathscr{L}_{i-1} - \mathscr{L}_{i-1}' \mathscr{L}_i}{\mathscr{L}_{i-1}^2} dy_i \\ &= \frac{1}{\mathscr{L}_{i-1}} \int \mathscr{L}_i' dy_i - \frac{\mathscr{L}_{i-1}'}{\mathscr{L}_{i-1}^2} \int \mathscr{L}_i dy_i \\ &= \frac{1}{\mathscr{L}_{i-1}} \int \mathscr{L}_i' dy_i - \frac{\mathscr{L}_{i-1}'}{\mathscr{L}_{i-1}} \\ &= 0 \end{aligned}$$

Thus

$$\Big\{\frac{\partial L_n(\theta)}{\partial\theta_a} = \sum_{i=1}^{n} \frac{\partial U_i(\theta)}{\partial\theta_a}, \mathscr{F}_n, n \geq 1\Big\}$$

is a martingale for $a = 1, ..., s$. Then, by WLLN, as $n \to \infty$:

$$\frac{L_n'(\theta)}{n} \xrightarrow{P} 0 \tag{4.3}$$

By(A3),(A4) and martingale CLT (Hall and Heyde,1980), we obtain

$$\frac{1}{\sqrt{n}}L_n'(\theta) \to N(0, \Gamma(\theta)) \tag{4.4}$$

for $\theta \in \Omega_0$.

On the other hand, for each element of matrix $L_n''(\theta_1)$, using Taylor's Expansion:

$$\frac{1}{n}\frac{\partial^2}{\partial\theta_a\partial\theta_b}L_n(\theta_1) = \frac{1}{n}\frac{\partial^2}{\partial\theta_a\partial\theta_b}L_n(\theta_0) + \frac{1}{n}\frac{\partial^3}{\partial\theta_a\partial\theta_b\partial\theta_c}L_n(\theta_2)(\theta_1 - \theta_0) \tag{4.5}$$

where θ_2 is a vector among the two balls with radii $||\theta_0||$ and $||\theta_1||$.

By (A6)

$$\frac{1}{n}\frac{\partial^2}{\partial\theta_a\partial\theta_b}L_n(\theta_0) \xrightarrow{P} -\gamma_{ab}(\theta_0)$$

By (A2)

$$\frac{1}{n}\left|\frac{\partial^3}{\partial\theta_a\partial\theta_b\partial\theta_c}L_n(\theta_2)\right| \leq \frac{1}{n}\sum_{i=1}^{n} M_i(Y_1, ..., Y_i) \leq M_{abc}$$

Thus,$\frac{1}{n}\frac{\partial^2}{\partial\theta_a\partial\theta_b}L_n(\theta_1)$ is bounded in probability if we let $||\theta_1 - \theta_0||$ be less than some constant. Therefore, in (4.4)

$$\hat{\theta}_n - \theta_0 \xrightarrow{P} 0 \tag{4.6}$$

Considering the consistence of $\hat{\theta}_n$, as $n \to \infty$, we have $\theta_1 - \theta_0 \xrightarrow{P} 0$. Therefore, the second item in (4.5) converges to 0 in probability. Then,

$$\frac{1}{n}\frac{\partial^2}{\partial\theta_a\partial\theta_b}L_n(\theta_1) \xrightarrow{P} -\gamma_{ab}(\theta_0)$$

Therefore,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\Big[\frac{L_n''(\theta_1)}{n}\Big]^{-1} \frac{L_n'(\theta_0)}{\sqrt{n}} \to N(0, \Gamma(\theta_0)^{-1})$$

The proof is complete.

Appendix 2

Proof of Theorem 2:

By the assumption, we have

$$\begin{aligned}
\overline{\mathbf{F}}_n &= \Big(\mathbf{I}+\frac{1}{n-1}\mathbf{H}\Big)\cdots\Big(\mathbf{I}+\frac{1}{i}\mathbf{H}\Big)\overline{\mathbf{F}}_i \\
&\quad + \sum_{j=i+1}^{n}\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{H}\Big)\cdots\Big(\mathbf{I}+\frac{1}{j}\mathbf{H}\Big)\boldsymbol{\Delta}\mathbf{W}_j \\
&= \mathbf{T}\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}'\Big)\cdots\Big(\mathbf{I}+\frac{1}{i}\mathbf{J}'\Big)\mathbf{T}^{-1}\overline{\mathbf{F}}_i \\
&\quad + \sum_{j=i+1}^{n}\mathbf{T}\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}'\Big)\cdots\Big(\mathbf{I}+\frac{1}{j}\mathbf{J}'\Big)\mathbf{T}^{-1}\boldsymbol{\Delta}\mathbf{W}_j
\end{aligned}$$

We consider the elements of the matrix

$$\begin{aligned}
&\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}^{'}\Big)\cdots\Big(\mathbf{I}+\frac{1}{j}\mathbf{J}^{'}\Big)\\
&=\begin{pmatrix}
\prod_{k=j}^{n-1}\Big(1+\frac{1}{k}\Big) & 0 & \cdots & 0\\
0 & \prod_{k=j}^{n-1}\Big(\mathbf{I}+k^{-1}\mathbf{J}_1^{'}\Big) & \cdots & 0\\
\cdots & \cdots & \cdots & \cdots\\
0 & 0 & \cdots & \prod_{k=j}^{n-1}\Big(\mathbf{I}+k^{-1}\mathbf{J}_r^{'}\Big)
\end{pmatrix}
\end{aligned}$$

By elementary calculus, we know that: as $n > j \to \infty$,

$$\Big(1+\frac{1}{n-1}\Big)\cdots\Big(1+\frac{1}{j}\Big)=\Big(\frac{n}{j}\Big)\Big(1+o(1)\Big)$$

and the $(h-k,h)$-element of the block matrix $\prod_{k=j}^{n-1}\Big(\mathbf{I}+k^{-1}\mathbf{J}_t^{'}\Big)\Big(\frac{j}{n}\Big)$ has the estimation

$$\frac{1}{k!}\Big(\frac{j}{n}\Big)^{1-Re(\lambda_t)}\log^k\Big(\frac{n}{j}\Big)(1+o(1))\le\frac{3}{k!}\Big(\frac{k}{e(1-|\lambda_t|-\epsilon)}\Big)^k\Big(\frac{j}{n}\Big)^{\epsilon} \tag{4.7}$$

where λ_t is the eigenvalue of J_t and $0<\epsilon<1-|\lambda_t|$. These imply that

$$\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}^{'}\Big)\cdots\Big(\mathbf{I}+\frac{1}{i}\mathbf{J}^{'}\Big)\Big(\frac{i}{n}\Big)\to\begin{pmatrix}1&0&\cdots&0\\0&0&\cdots&0\\\cdots&\cdots&\cdots&\cdots\\0&0&\cdots&0\end{pmatrix}=e_1^{'}e_1 \tag{4.8}$$

where $e_1=(1,0,...,0)$.

$$
\begin{aligned}
&\Delta W_{k,\ j} \\
= &\left[\frac{(Kp_{kj}-1)F_{k,\ j-1}(x_j)+\sum_l q_{lj}F_{l,\ j-1}(x_j)}{(K-1)(j-1)} - \frac{\overline{F}_{k,\ j-1}\int_{x_j}(Kp_{kj}-1)+\sum_l \overline{F}_{l,\ j-1}\int_{x_j} q_{lj}}{(K-1)(j-1)}\right] \\
- &\frac{K}{(K-1)(j-1)[K+(j-1)g(x_j)]}[(Kp_{kj}-1)F_{k,\ j-1}(x_j)+\sum_l q_{lj}F_{l,\ j-1}(x_j)] \\
+&\frac{g(x_j)}{K-1}\left[\frac{1}{K+\sum_{u=1}^{j-1}f(x_j,x_u)}-\frac{1}{K+(j-1)g(x_j)}\right][(Kp_{kj}-1)F_{k,\ j-1}(x_j)+\sum_l q_{lj}F_{l,\ j-1}(x_j)] \\
+ &\int_x \left[A_{k,\ j}-E(A_{k,\ j}\Big|\mathscr{F}_{j-1})\right]
\end{aligned}
$$

We are going to show that as $i \to \infty$,

$$
n^{-1}\sum_{j=i+1}^{n}\mathbf{T}\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}'\Big)\cdots\Big(\mathbf{I}+\frac{1}{j}\mathbf{J}'\Big)\mathbf{T}^{-1}\mathbf{\Delta W}_j \overset{a.s.}{\to} 0 \tag{4.9}
$$

Note that the first row of T^{-1} is $(1,1,...,1)$, as $n > j \to \infty$,

$$
\begin{aligned}
\Big(\frac{n}{j}\Big)\mathbf{T}\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}'\Big)\cdots\Big(\mathbf{I}+\frac{1}{j}\mathbf{J}'\Big)\mathbf{T}^{-1} &\to \mathbf{T}e_1{}'e_1\mathbf{T}^{-1} \\
&= \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}(1,1,...,1)
\end{aligned}
$$

Thus,

$$
e_1\mathbf{T}^{-1} = (1,1,...,1)\Delta\mathbf{W}_j = \sum_k \Delta W_{k,\ j}
$$

Suppose f is bounded by $0 < m < f < M, (m, M$ are constants), then the second term of $\Delta W_{k,\ j}$ has order $O(j^{-1})$, the third term is bounded by the martin-

gale $\frac{1}{j-1}\sum_{u=1}^{j-1}[f(x_j,x_u)-g(x_j)]$, then has order $O(j^{-1/2})$. The summation of the first term of $\Delta W_{k,\ j}$ under k,is $\frac{1}{j-1}\sum_{u=1}^{j-1}[f(x_j,x_u)-g(x_u)]$, has order $O(j^{-c})$ for some constant $c>0$ under a certain assumption on $g(x)$. The summation of the last term under k, is 0.

Thus, $(1,1,...,1)\Delta\mathbf{W}_{j,\ k}$ has order $O(j^{-\delta})$ for some $\delta>0$, therefore,

$$n^{-1}\sum_{j=i+1}^{n}\mathbf{T}\Big(\mathbf{I}+\frac{1}{n-1}\mathbf{J}^{'}\Big)\cdots\Big(\mathbf{I}+\frac{1}{j}\mathbf{J}^{'}\Big)\mathbf{T}^{-1}\mathbf{\Delta W}_j=\sum_{j=i+1}^{n}j^{-1}\begin{pmatrix}1\\1\\\cdots\\1\end{pmatrix}(1+o(1))\sum_k\Delta W_{k,\ j}$$

and (4.9) follows.

This leads to:

$$T^{-1}\overline{\mathbf{F}}_n/n-e_1^{'}e_1T^{-1}\overline{\mathbf{F}}_i/i\to 0$$

as $n>i\to\infty$. Since each element of $T^{-1}\overline{\mathbf{F}}_n/n$ is bounded, we conclude that $T^{-1}\overline{\mathbf{F}}_n/n$ must converge to a limit, say z, satisfying $z=e_1^{'}e_1z$. This implies that $z\propto e_1^{'}$. It follows that

$$\overline{\mathbf{F}}_n/n\to CTe_1^{'}=C\mathbf{v}$$

C is a constant.By(2.11),$C=g$.

Therefore

$$
\begin{aligned}
F_{k,\ n}(x) &= 1+\delta_0+\sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\left[\frac{(Kp_{ki}-1)F_{k,\ i-1}(x_i)}{i-1}+\sum_l\frac{q_{li}F_{l,\ i-1}(x_i)}{i-1}\right]\\
&- \sum_{i=2}^{n}\frac{Kf(x,x_i)}{[K+(i-1)g(x_i)]g(x_i)}\left[\frac{(Kp_{ki}-1)F_{k,\ i-1}(x_i)}{i-1}+\sum_l\frac{q_{li}F_{l\ i-1}(x_i)}{i-1}\right]\\
&+ \delta_{1n}+\delta_{2n} \qquad\qquad (\delta_{2n}\equiv\sum_{i=1}^{n}\left[A_{k,\ i}-E\big(A_{k,\ i}\big|\mathscr{F}_{i-1}\big)\right])\\
&= 1+\delta_0+\sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\left[(Kp_{ki}-1)g(x_i)v_k+\sum_l q_{li}g(x_i)v_l\right]\\
&+\sum_{i=1}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\left[(Kp_{ki}-1)\Big(\frac{F_{k,\ i-1}(x_i)}{i-1}-g(x_i)v_k\Big)+\sum_l q_{li}\Big(\frac{F_{l,\ i-1}(x_i)}{i-1}-g(x_i)v_l\Big)\right]\\
&- \sum_{i=2}^{n}\frac{Kf(x,x_i)}{[K+(i-1)g(x_i)]g(x_i)}\left[\frac{(Kp_{ki}-1)F_{k,\ i-1}(x_i)}{i-1}+\sum_l\frac{q_{li}F_{l,\ i-1}(x_i)}{i-1}\right]\\
&+ \delta_{1n}+\delta_{2n}
\end{aligned}
$$

Using the assumption about the covariance between f and p_{ki}, we have:

$$
\begin{aligned}
& n^{-1}\sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\left[(Kp_{ki}-1)g(x_i)v_k+\sum_l q_{li}g(x_i)v_l\right]\\
=\ & n^{-1}\sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)}\left[(Kp_{ki}-1)v_k+\sum_l q_{li}v_l\right]\\
\overset{a.s.}{\to}\ & (K-1)^{-1}E_{x_i}f(x,x_i)\left[(Kp_{ki}-1)v_k+\sum_l q_{li}v_l\right]\\
=\ & (K-1)^{-1}g(x)\left[(KP_k-1)v_k+\sum_l Q_l v_l\right]\\
=\ & g(x)v_k
\end{aligned}
$$

The term:

$$
\sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\left[(Kp_{ki}-1)\Big(\frac{F_{k,\ i-1}(x_i)}{i-1}-g(x_i)v_k\Big)+\sum_l q_{li}\Big(\frac{F_{l,\ i-1}(x_i)}{i-1}-g(x_i)v_l\Big)\right]
$$

can be written as:

$$\sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\Big[(Kp_{ki}-1)\big(\frac{F_{k,\ i-1}(x_i)}{i-1}-\frac{\overline{F}_{k,\ i-1}}{i-1}\big)+\sum_l q_{li}\big(\frac{F_{l,\ i-1}(x_i)}{i-1}-\frac{\overline{F}_{l,\ i-1}}{i-1}\big)\Big]$$
$$+\ \sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\big[(Kp_{ki}-1)\big(\frac{\overline{F}_{k,\ i-1}}{i-1}-gv_k\big)+\sum_l q_{li}\big(\frac{\overline{F}_{l,\ i-1}}{i-1}-gv_l\big)\big]$$
$$+\ \sum_{i=2}^{n}\frac{f(x,x_i)}{(K-1)g(x_i)}\big[(Kp_{ki}-1)\big(gv_k-g(x_i)v_k\big)+\sum_l q_{li}\big(gv_l-g(x_i)v_l\big)\big]$$

Since $\big\{\frac{F_{k,\ i-1}(x_i)}{i-1}-\frac{\overline{F}_{k,\ i-1}}{i-1}, i=2,...,n\big\}$ is a martingale difference, and $\frac{f(x,x_i)}{g(x_i)}$ is bounded, the first term is therefore $o(n)$, the second, by LLN, is $o(n)$, the third term is also $o(n)$. $n^{-1}\sum_{i=2}^{n}\frac{Kf(x,x_i)}{[K+(i-1)g(x_i)]g(x_i)}\big[\frac{(Kp_{ki}-1)F_{k,\ i-1}(x_i)}{i-1}+\sum_l\frac{q_{li}F_{l\ i-1}(x_i)}{i-1}\big]$ is also converges to 0. $\delta_{1n}/n, \delta_{2n}/n$ automatically converge to 0.

Therefore,

$$\frac{F_{k,\ n}(x)}{n}\overset{a.s.}{\to} g(x)v_k$$

and

$$\frac{F_{k,\ n}(x)}{\sum_{l=1}^{K}F_{l,\ n}(x)}\overset{a.s.}{\to} v_k$$

for $k=1,...,K$.

By Stolz Therorem,

$$n^{-1}\sum_{i=1}^{n}I_{k,\ i}=n^{-1}\sum_{i=1}^{n}\frac{F_{k,\ i-1}(x_i)}{\sum_{l=1}^{K}F_{l,\ i-1}(x_i)}$$

has the same limit as $\frac{F_{k,\ n}(x_i)}{\sum_{l=1}^{K}F_{l,\ n}(x_i)}$.

The proof is complete.

Appendix 3

Simulation Source Code: A Typical Example

```
options(expressions=100000)

N<-400 #No. of Circles


n<-1000 #No.of Patients in Each Circle


a<--1/2 b<-10 c<--10 e<-2 #Set True Value for Parameters

A<-array(1,dim=c(n,4,N)) #Covariates Array


y<-array(0,dim=c(n,N)) #response array


res<-array(0,dim=c(N,4)) #parameter estimation


for(j in 1:N) {
```

```
x<-rnorm(n,mean=0,sd=1)  #X from Standard Normal Distribution
#x<-runif(n, min=0,max=1)
  A[,4,j]<-x
  I<-matrix(c(rep(0,10*n)),ncol=n,byrow=T) #Allcation Array
  P<-I #Probability of Success Array
  af<-function(k,i,z){     #Define Allocation Rule Function
  if(i==0) aff<-1
  else aff<-af(k,i-1,z)+(y[i,j]*I[k,i]+(1-y[i,j])*(1-I[k,i])/9)
  /(1+(abs(z-x[i])))
  return(aff)}
  d<-c(rep(0,n)) # Dose Choice
  d[1]<-sample(10,1)
  A[1,2,j]<-d[1]/10
  A[1,3,j]<-d[1]^2/100
  I[d[1],1]<-1
  P[d[1],1]<-exp(a+b*d[1]/10+c*d[1]^2/100+e*x[1])
  /(1+exp(a+b*d[1]/10+c*d[1]^2/100+e*x[1]))
  y[1,j]<-sample(0:1,1,T,c(1-P[d[1],1],P[d[1],1]))
```

```
for(i in 2:n) {   urn<-rep(0,10)
   # f<-rep(1,10)
   # for(l in 1:(i-1))
   # {for(k in 1:10)
   # {f[k]<-f[k]+(y[l,j]*I[k,l]+(1-y[l,j])*(1-I[k,l])/9)/(1+abs(x[i]-x[l]))}
#} # Simplify Allocation Proportional Function to Avoid Stack Flow
   # urn<-f
    urn<-af(1:10,i-1,x[i])
    urn<-urn/sum(urn)
d[i]<-sample(1:10,1,T,urn)
    I[d[i],i]<-1
    P[d[i],i]<-exp(a+b*d[i]/10+c*d[i]^2/100+e*x[i])
    /(1+exp(a+b*d[i]/10+c*d[i]^2/100+e*x[i]))
    A[i,2,j]<-d[i]/10
    A[i,3,j]<-d[i]^2/100
   y[i,j]<-sample(0:1,1,T,c(1-P[d[i],i],P[d[i],i]))

}
```

```
 par<-glm(formula=cbind(y[,j],1-y[,j])~A[,2,j]+A[,3,j]+A[,4,j]
 ,fam=binomial,maxit=100) #MLE


 for(t in 1:4)
       {
    res[j,t]<-par$coefficients[t]
} }


###Maximum Likelihood Estimation######
me<-apply(res,2,mean)
me
var(res)


shapiro.test(res[,1]) #Normality TEST
shapiro.test(res[,2])
shapiro.test(res[,3])
shapiro.test(res[,4])


###finding optimal dose#####
w<-10000
y<-rep(0,w)
```

```
a<--1/2 b<-10c<--10 e<-2


d<-c(rep(0,w)) I<-matrix(rep(0,10*w),ncol=w,byrow=T)


x<-rnorm(w,mean=0,sd=1) d[1]<-sample(10,1)


pro<-array(0,dim=c(w,2))
pro[1,1]<-exp(a+b*d[1]/10+c*d[1]^2/100+e*x[1])
/(1+exp(a+b*d[1]/10+c*d[1]^2/100+e*x[1]))
pro[1,2]<-d[1]
  y[1]<-sample(0:1,1,T,c(1-pro[1,1],pro[1,1]))


for(i in 2:w) {   urn<-rep(0,10) f<-rep(1,10)
  for(j in 1:(i-1))
    {for(k in 1:10)
    {f[k]<-f[k]+(y[j]*I[k,j]+(1-y[j])*(1-I[k,j])/9)/(1+abs(x[i]-x[j]))}
   }
    urn<-f
    urn<-urn/sum(urn)
  d[i]<-sample(1:10,1,T,urn)
```

```
    I[d[i],i]<-1
    pro[i,1]<-exp(a+b*d[i]/10+c*d[i]^2/100+e*x[i])
    /(1+exp(a+b*d[i]/10+c*d[i]^2/100+e*x[i]))
    pro[i,2]<-d[i]
   y[i]<-sample(0:1,1,T,c(1-pro[i,1],pro[i,1]))
    }



pp<-matrix(rep(0,30),ncol=10,byrow=T)
for (k in 1:10) {
 for (i in 1:w) { if(pro[i,2]==k)
     {pp[1,k]<-k/10 pp[2,k]<-pp[2,k]+pro[i,1] pp[3,k]<-pp[3,k]+1} }
pp[2,k]<-pp[2,k]/pp[3,k] }
pp


par(mfrow=c(1,1))
x1<-seq(0.1,1,0.05)
y1<-a+b*x1+c*x1^2
y1<-exp(y1)/(1+exp(y1))
plot(pp[1,],pp[2,],ylim=c(0.3,0.9))
lines(x1,y1)
```

```
rr<-rep((1-pp[2,]),10)

h<-matrix(rr,ncol=10,byrow=T)

h<-h/9

diag(h)<-pp[2,]


g<-function(z)

{ga<-0 for(i in 1:w)

{ ga<-ga+1/(1+(abs(z-x[i]))) }

ga<-ga/w return(ga) }

ev<-eigen(h)$vectors[,1]/sum(eigen(h)$vectors[,1])


ev

f/sum(f)

f/(w-1)

ev*g(x[w])
```

Bibliography

[1] Athreya, K. B. (1967). Limit Theorem for Multitype Contuinuous Time Markov Branching Processes and Some Classical Urn Schemes. *Standford: Stanford University(Doctoral Thesis).*

[2] Athreya, K.B. and Karlin, S. (1967) Limit Theorems for the Split Times of Branching Processes. *Journal of Mathematics and Mechanicss* **17** 257-277.

[3] Athreya, K.B and Karlin, S. (1968). BEmbedding of Urn Schemes into Contunious Time Markov Branching Processes and Related limit Theorems. *Annals of Mathematics Statistics* **39** 1801-1817 .

[4] Athreya, K.B. and Ney, P. (1972). Branching Processes. *Berlin:Springer.*

[5] Bai, Z.D and Hu, F. (1999). Asymptotic Theorems for Urn Models with Nonhomogeneous Generating Matricies. *Stochastic Processes and Their Applications* **80** 87-101.

[6] Bai, Z.D., Hu, F.,and Rosenberger, W.F. (2002).Asymptotic Properties of Adaptive Designs for Clincal Trials with Delayed Response. *Annals of Statistics* **30, No 1**,1-18.

[7] Bickel, P.J. and Doksum, K.A. (2001). Mathematical Statistics (Basic Ideas and Selected Topics), 2nd ed. *Prentice Hall* **Vol 1**.

[8] Billingsley, P. (1995). Probability and Measure, 3rd ed. *John Wiley & Sons*.

[9] Hall, P. and Heyde, C.C. (1980). Martingale Limit Theory and Its Application. *London:Academic Press*.

[10] Johnson, R.A. and Wichern, D.W. (2002). Applied Multivariate Statistical Analysis, 5th ed. *Prentice Hall*.

[11] Rosenberger, W. F. (1996). New Directions in Adaptive Designs. *Statistical Science* 11, 137-149.

[12] Rosenberger, W. F. (2002). Randomized Urn Models and Sequential Design. *Sequential Analysis (with discussion)* **21**, 1-41.

[13] Rosenberger, W. F., Flournoy, N., and Durham, S. D. (1997). Asymptotic Normality of Maximum Likelihood Estimators from Multiparameter Response-Driven Designs. *Journal of Statistical Planning and Inference* 60, 69-76.

[14] Rosenberger, W. F. and Hu, M. (2002). On the Use of Generalized Linear Models Following a Sequential Design. *Statistics and Probability Letters* **56**, 155-161.

[15] Rosenberger, W.F. and Sriram, T.N. (1997). Estimation for an Adaptive Allocation Design. *Journal of Statistical Planning and Inference* **59** 309-319.

[16] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples).*Biometrika* **52** 3 and 4, pages 591-611..

[17] Smythe, R.T. (1996) Central Limit Theorems for Urn Models. *Stochastic Processes and Their Applications* **65** 115-137.

[18] Xu, J. (2003). Maximum Likelihood Estimators from a New Adaptive Design Association with Covariates. *Singapore:National University of Singapore (Master Thesis).*

[19] Veneble, W.N. and Ripley, B.D. (2002). Morden Applied Statistics with S, 4t ed. *New York:Springer-Verlag.*

[20] Wei, L.J. and Durham, S. (1978). The Randomized Play-the-Winner Rule in Medical Trials. *Journal of the American Statistical Association* **73** 840-843.

[21] Zelen, M. (1969). Play the Winner Rule and the Controlled Clinical Trial. *Journal of the American Statistical Association* **64** 131-146.