ASYMPTOTIC PROPERTIES OF COVARIATE-ADJUSTED
ADAPTIVE DESIGNS

BY LI-XIN ZHANG1, FEIFANG HU2,∗, SIU HUNG CHEUNG3 AND WAI SUM CHAN3

Zhejiang University, University of Virginia, Chinese University of Hong Kong and
University of Hong Kong

Owing to the benefits that higher proportions of patients are likely to receive
the better treatment, response-adaptive designs are considered to be valuable
statistical tools in clinical studies. Nevertheless, there is a lack of a com-pre-
hensive study of adaptive designs with the inclusion of covariates, regardless of
their importance in clinical experiments. The major reason is that covariate-
adjusted adaptive designs are extreme complex to formulate. For covariate-
adjusted adaptive designs, both the allocation scheme and the estimation of
parameters are being affected not only by the responses, but also by the co-
variates. In this paper, we have overcome the technical hurdles and layout the
framework of the general covariate-adjusted adaptive design for the allocation
of subjects to K (≥ 2) treatments. This design can be applied to different types
of responses. The asymptotic properties and the advantages of this general
adaptive design are studied under some widely satisfied conditions. Two im-
portant special cases, linear model and logistic regression model, are considered
in details. For comparing two treatments with binary responses, we find that
the covariate-adjusted design allocates more subjects to the better treatment
at each given covariate level, and the overall success proportion is higher than
treatment allocation schemes without incorporating covariates.

Abbreviated Title: Covariate-adjusted adaptive design.
1. Introduction. In most clinical trials, patients accrue sequentially. Adaptive designs provide efficient allocation schemes which assign different treatments to incoming patients based on previous observed patients’ responses. The most valuable contribution of adaptive design in clinical studies is its promising capability to minimize the number of patients assigned to the inferior treatment, while simultaneously, permitting useful statistical inferences. The ethical character of adaptive designs can be illustrated by the clinical study of the risk of maternal-to-infant HIV transmission (Zelen and Wei (1995)). For this study, a 50-50 allocation scheme is used to assign treatments (AZT and placebo) to patients. If the treatment assignment had been implemented with a “play-the-winner rule” of the adaptive allocation scheme (Wei and Durham, 1978), only 60 infants instead of 80 would have been born with HIV positive. With the adaptive allocation scheme, patient responses are continually being evaluated to revise the probabilities of treatment allocation proportions. Therefore, more pregnant women would have received the better treatment (AZT) during the experiment.

Early important works in adaptive designs can be attributed to Thompson (1933) and Robbins (1952). Since then, a steady production of research works in this area has generated various treatment allocation schemes in clinical applications. To classify historical developments, a simplified approach is to group previous works into two different categories (Rosenberger and Lachin, 2002): (1) urn models, and (2) sequential estimation procedures.

Urn models were early proposed and studied by Athreya and Karlin (1968), Zelen (1969), Wei and Durham (1978) and others. For Urn models, there are various types of balls, each represents a particular treatment. At each stage of treatment allocation, the allocation probabilities depend on the numbers of various types of balls. Response of each patient after treatment plays an essential role in the determination of subsequent urn compositions (the numbers of balls in the urn). The basic strategy is to “reward” more balls to successful treatments. One can refer to a discussion paper by Rosenberger (2002) for recent developments.

Sequential estimation procedure is another important class of adaptive designs. One major difference between urn models and sequential estimation procedures is
that for the latter, pre-specified targets can be incorporated in the derivation of the design. A sequential estimation scheme is a crucial ingredient of the design. Research works in this area are abundant. An influential sequential estimation procedure, the doubly-adaptive biased coin design (DBCD), was proposed by Eisele (1994), and Eisele and Woodroofe (1995). Recently, Hu and Zhang (2004) proposed a family of doubly-adaptive biased coin designs and studied its asymptotic properties.

To compare various adaptive designs, an extensive investigation was launched by Hu and Rosenberger (2003). For comparing efficacies of two treatments with binary responses, Rosenberger and Hu (2004) demonstrate with a simulation study that the doubly adaptive biased coin design is always as powerful or slightly more powerful than complete randomization, with fewer expected treatment failures.

In clinical trials, covariate information is usually available and often has a strong influence on patient's response. For instance, the efficacy of a hypertensive drug is related to a patient’s initial blood pressure and cholesterol level. The effectiveness of a cancer treatment may depend on whether the patient is a smoker or a non-smoker. Male and Female subjects can react very differently to a certain pain reliever. In these circumstances, adaptive designs based solely on the patient’s responses are not appropriate.

Even though the literature is vast in adaptive designs, the effort to incorporate covariate information is still in its infancy. Without the covariates, the formulation of the allocation scheme and the estimation of parameters in the model rely only on the responses of observed patients. However, with the inclusion of covariates, both the allocation scheme and the estimation of parameters are being affected not only by the responses, but also by the covariates. That generates a certain level of technical complexity which hinders the advancement of the theoretical development of adaptive designs with covariates.

Therefore, the mission of this paper is to provide a more comprehensive theoretical foundation for adaptive designs with covariates. Practitioners in clinical studies will find our derived allocation scheme, together with the proved statistical properties a useful paradigm for treatment allocation in trials where covariates are important.

As mentioned, the history of incorporating covariates in adaptive designs is short.
A more recent attempt can be found in Rosenberger, Vidyashankar and Agarwal (2001) which considered a covariate-adjusted response-adaptive design for binary responses. Their encouraging simulation study indicates that the inclusion of the covariate reduces significantly the percentage of treatment failures, bearing substantial benefits in terms of ethical costs. However, theoretical justifications and asymptotic properties are not given. Further, the applications of their procedure are limited to two treatments with binary responses. Another interesting article by Bandyopadhyay and Biwas (2001) considered a linear model to incorporate covariate information under the assumption of normal errors. However, the allocation scheme is not covariate adjusted and their method only applies to normal responses.

To summarize, the major task of this paper is to provide a more comprehensive study of covariate-adjusted adaptive designs. The main objectives of this paper are (i) to propose a general covariate-adjusted adaptive design that can be applied to cases where $K$-treatments ($K \geq 2$) are present and to different types of responses (discrete or continuous), and (ii) to study important asymptotic properties of the covariate-adjusted adaptive design. Major mathematical techniques including martingale theory and Gaussian approximation are employed in order to tackle the mathematical complexity of the formulation of the allocation scheme.

The paper is organized as follows. Before the introduction of a more general framework, we start with the simplest case of the covariate-adjusted adaptive design for illustrative purposes. Therefore, in Section 2, the two-treatment clinical trials with dichotomous responses and logistic regression are being considered. The outline of our proposed covariate-adjusted adaptive design is given. Afterwards, we derive the asymptotic properties of the average allocation proportion to each treatment and the allocation proportions for a given covariate. In clinical trials, the evaluation of the expected number of treatment successes is in general essential. In the absence of any covariate, the expected number of treatment successes can be directly computed from the allocation proportion. However, it is far more complicated for covariate-adjusted adaptive design. The properties of the expected total number of treatment successes and the success proportion of the covariate-adjusted adaptive design are extensively examined.
To demonstrate the usefulness of our proposed design, we compare covariate-adjusted adaptive designs with ordinary adaptive designs that ignore the available covariate information. As expected, the covariate-adjusted design is found to assign more subjects to the better treatment for each given covariate. In addition, the covariate-adjusted adaptive design yields more expected treatment successes.

In Section 3, the general covariate-adjusted adaptive design is provided, which can be applied to (i) different types of responses and (ii) trials with more than two treatments. Useful asymptotic results including the strong consistencies and asymptotic normalities of the estimators of unknown parameters for the general covariate-adjusted adaptive design are given. Further, we establish the asymptotic properties of the average allocation proportions, the allocation proportions given the covariate and the success proportions, for more general models. In Section 4, the general covariate-adjusted adaptive design is applied to the linear regression model. Then, the popular logistic regression model for general k-treatment trials are considered in Section 5. Some further discussions are given in Section 6. Finally, technical proofs are outlined in Appendix.

2. Comparing two treatments for binary responses. Consider two treatments, A and B, with binary responses (success and failure). Let $\eta$ be a $d$-vector of covariates. For a given covariate $\eta$, let $p_A = p_A(\theta_A, \eta)$ and $p_B = p_B(\theta_B, \eta)$ be the probability of success for a subject on treatment A and B, respectively. Write $q_A = 1 - p_A$ and $q_B = 1 - p_B$. We assume

$$\logit(p_A) = \alpha_A + \theta_A \eta' \quad \text{and} \quad \logit(p_B) = \alpha_B + \theta_B \eta'$$

where $(\alpha_A, \theta_A)$ and $(\alpha_B, \theta_B)$ are regression parameters. Here $\theta_A$ and $\theta_B$ are $d$-dimensional vectors. Without the lose of generality, we assume $\alpha_A = \alpha_B = 0$, or otherwise, we can redefine the covariate vector to be $(1, \eta)$.

When patients are arrive sequentially, the allocation rule is defined as the following. First, the first $m_0$ subjects are randomized to treatments A or B with probability $1/2$ each. For the $(m + 1)$th patient $(m \geq m_0)$, the assignment of treatment will be based on the responses of previous $m$ patients. The maximum likelihood estimators, $\hat{\theta}_{m,A}$ of $\theta_A$ and $\hat{\theta}_{m,B}$ of $\theta_B$ are evaluated by fitting the logistic model. Let $X_i = (X_{i,A}, X_{i,B})$
denote a treatment indicator of the $i$-th randomization ($(1, 0)$ if $A$; $(0, 1)$ if $B$), and $\eta_i$ be the covariate of the $i$-th subject. Let $\mathcal{F}_m$ be the $\sigma$-field based on the history of treatments assignments, responses, and covariates for subjects $1, \ldots, m$. Now assume the $(m + 1)$-th subject is ready to be randomized and that its covariate $\eta_{m+1}$.

We allocate this person to treatments $A$ or $B$ with a probability proportional to the estimated covariate-adjusted odds ratio, comparing treatments $A$ and $B$, evaluated at $\eta_{m+1}$:

$$p_A(\hat{\theta}_{m,A}, \eta_{m+1})/q_A(\hat{\theta}_{m,B}, \eta_{m+1}) = \frac{\exp\{\hat{\theta}_{m,A}'\eta_{m+1}\}}{\exp\{\hat{\theta}_{m,B}'\eta_{m+1}\}}.$$ 

That is, we allocate the $(m + 1)$-th subject to treatment $A$ with a probability

$$P(X_{m+1,A} = 1 | \mathcal{F}_m, \eta_{m+1}) = \frac{f_A(\hat{\theta}_{m,A}, \hat{\theta}_{m,B}, \eta_{m+1})}{\exp\{\hat{\theta}_{m,A}'\eta_{m+1}\} + \exp\{\hat{\theta}_{m,B}'\eta_{m+1}\}.$$ 

Here

$$f_A(\theta, \eta) = \frac{p_A/\rho_A + p_B/\rho_B}{\exp\{\theta_A'\eta\} + \exp\{\theta_B'\eta\}}$$

is a function of $\theta = (\theta_A, \theta_B)$ and the covariate $\eta$. The allocation rule in (2.2) is used by Rosenberger, Vidyashankar and Agarwal (2001). One can define other allocation rules by choosing the function $f_A$ suitably. For example, if one wishes to allocate subjects with probabilities proportional to the square root of the estimated success rates, then $f_A(p_A, p_B) = \sqrt{p_A}/(\sqrt{p_A} + \sqrt{p_B})$. Further discussions will be given in Section 6.

Let $N_{n,A} = \sum_{m=1}^n X_{m,A}$ be the number of subjects assigned to treatment $A$ in $n$ trials. It is frequent in clinical studies that the covariates $\eta$ are discrete variables. In these cases, we are able to derive the conditional allocation proportions. Let $N_{n,A|y} = \sum_{m=1}^n X_{m,A}I\{\eta_m = y\}$ be the number of subjects assigned to treatment $A$ in $n$ trials for a given covariate vector $y$ and $N_n(y) = \sum_{m=1}^n I\{\eta_m = y\}$ be the total number of subjects with covariate $y$. We have the following theorem.

**Theorem 2.1** Let $f_B = f_B(\theta_A, \theta_B, \eta) = 1 - f_A$, $I_A = E[f_A p_A q_A' \eta']$ and $I_B = E[f_B p_B q_B' \eta']$. Assume (i) $E[\eta^2 + \delta] < \infty$ for some $\delta > 0$; (ii) $E[\eta \eta']$ is a positive definite matrix; (iii) The parameter spaces of $\theta_A$ and $\theta_B$ are bounded, and the true values of parameters are their interior points. Then
(a) The maximum likelihood estimators $\hat{\theta}_{n,A}$ and $\hat{\theta}_{n,B}$ are strong consistent and

$$\sqrt{n}(\hat{\theta}_{n,A} - \theta_A, \hat{\theta}_{n,B} - \theta_B) \xrightarrow{D} N(0, \text{diag}(I_A^{-1}, I_B^{-1})).$$

(b) For the allocation proportion $N_{n,A}/n$, we have

$$N_{n,A}/n \rightarrow v_A =: E[f_A] \quad \text{a.s.} \quad \text{and} \quad \sqrt{n}(N_{n,A}/n - v_A) \xrightarrow{D} N(0, \sigma^2_A),$$

where

$$\sigma^2_A = v_A(1 - v_A) + 2 \left[ E\left[\frac{\partial f_A}{\partial \theta_A}\right] I_A^{-1}\left(E\left[\frac{\partial f_A}{\partial \theta_A}\right]\right) + E\left[\frac{\partial f_A}{\partial \theta_B}\right] I_B^{-1}\left(E\left[\frac{\partial f_A}{\partial \theta_B}\right]\right)\right],$$

(c) Given a covariate $y$ with $P(\eta = y) > 0$, we have

$$N_{n,A|y}/N_n(y) \rightarrow v_{A|y} \quad \text{a.s.} \quad \text{and} \quad \sqrt{N_n(y)/(N_{n,A|y}/N_n(y) - v_{A|y})} \xrightarrow{D} N(0, \sigma^2_{A|y}),$$

where $v_{A|y} = f_A(\theta_A, \theta_B, y)$ and

$$\sigma^2_{A|y} = v_{A|y}(1 - v_{A|y})$$

$$+ 2 \left[ \frac{\partial f_A(\theta, y)}{\partial \theta_A} I_A^{-1}\left(\frac{\partial f_A(\theta, y)}{\partial \theta_A}\right) + E\left[\frac{\partial f_A(\theta_A, y)}{\partial \theta_B}\right] I_B^{-1}\left(\frac{\partial f_A(\theta, y)}{\partial \theta_B}\right)\right] P(\eta = y).$$

When the covariate $\eta$ is a discrete vector, by Theorem 2.1 (c), for each given covariate vector $y$ the allocation rule defined in (2.2) assigns more subjects to the better treatment. The proportion $N_{n,A|y}/N_n(y)$ can be conceptualized as the allocation proportion for a given level of the covariate $\eta = y$.

However, when $\eta$ is a continuous vector, Theorem 2.1 (c) is no longer valid. Alternatively, one may define the conditional proportion as

$$\frac{\sum_{m=1}^n X_{m,A} I\{\eta_m \in U(y, h_n)\}}{\sum_{m=1}^n I\{\eta_m \in U(y, h_n)\}},$$

where $U(y, h_n)$ is a ball with radius $h_n$ in $\mathbb{R}^d$. Also it can be showed that this conditional proportion is convergent to $f_A(\theta, y)$ in probability, whenever $\eta$ has a
density function and \( h_n \rightarrow 0, \, nh_n^d \rightarrow \infty \). However, the convergent rate may be unacceptably slow. So, when the covariate is a continuous vector (even when it is discrete but has many of states), the conditional proportion can only serve as a rough reference statistic.

Fortunately, in clinical trials, we are usually interested in the more informative statistic, the proportion of successes in \( n \) trials, \( S_n/n \), where \( S_n \) is the total number of successes in \( n \) trials. The following theorem provides the useful asymptotic properties of the proportion \( S_n/n \).

**Theorem 2.2** Suppose the conditions in Theorem 2.1 are satisfied. Then

\[
S_n/n \to \mu_S \text{ a.s. and } \sqrt{n}(S_n/n - \mu_S) \xrightarrow{D} N(0, \sigma^2_S),
\]

where \( \mu_S = E[f_Ap_A] + E[f_Bp_B] \) and

\[
\sigma^2_S = \mu_S(1 - \mu_S) + 2 \sum_{j,k=A}^B E[f_jp_jq_j]I_j^{-1}E[p_k\frac{\partial f_k}{\partial \theta_j}]'
+ 2 \sum_{j=A}^B \left( \sum_{k=A}^B E[p_k\frac{\partial f_k}{\partial \theta_j}] \right) I_j^{-1} \left( \sum_{k=A}^B E[p_k\frac{\partial f_k}{\partial \theta_j}] \right)'
= \mu_S(1 - \mu_S) + 2(E[f_Ap_Aq_A]I_A^{-1} - E[f_Bp_Bq_B]I_B^{-1})E[f_Af_B(p_A - p_B)\eta]'
+ 2E[f_Af_B(p_A - p_B)\eta](I_A^{-1} + I_B^{-1})E[f_Af_B(p_A - p_B)\eta]'.
\]

To evaluate the advantage of our proposed procedure, we will compare the covariate-adjusted design with the corresponding design which ignores the covariate information. Before launching the comparative study, we need to give corresponding asymptotic properties for the latter. Let \( p_A^* = E[p_A(\theta_A, \eta)] \) and \( p_B^* = E[p_B(\theta_B, \eta)] \) be the average success probabilities of a subject on treatments \( A \) and \( B \), respectively. Write \( q_A^* = 1 - p_A^* \) and \( q_B^* = 1 - p_B^* \). Here, we describe the allocation procedure: To start, we randomize the first \( m_0 \) subjects to treatments \( A \) or \( B \) with probability 1/2 each. For the \((m + 1)\)th \((m \geq m_0)\) subject, the assignment of treatments is done according to the maximum likelihood estimators \( \hat{p}_{m,B}^* \) of \( p_A^* \) and \( p_B^* \) respectively, based on the responses of previous \( m \) subjects. Let \( \hat{q}_{m,A}^* = 1 - \hat{p}_{m,A}^* \) and \( \hat{q}_{m,B}^* = 1 - \hat{p}_{m,B}^* \). Now, the \((m + 1)\)-th subject is assigned to treatment \( A \) with probability:

\[
P(X_{m+1,A} = 1|F_m) = f_A^*(\hat{p}_{m,A}^*, \hat{p}_{m,B}^*) = \frac{\hat{p}_{m,A}/\hat{q}_{m,A}^*}{\hat{p}_{m,A}/\hat{q}_{m,A}^* + \hat{p}_{m,B}/\hat{q}_{m,B}^*},
\]
where \( f_A^* = f_A^*(p_A^*, q_A^*) = \frac{p_A^*/q_A^*}{p_A^*/q_A^* + p_B^*/q_B} \) is a function of \( p_A^* \) and \( p_B^* \).

**Proposition 2.1** Let \( f_B^* = 1 - f_A^* \), \( v_A^* = f_A^* \) and

\[
\sigma_{MLE}^2 = \left( \frac{\partial f_A^*}{\partial p_A^*} \right)^2 + \left( \frac{\partial f_B^*}{\partial p_B^*} \right)^2 = (f_A^*)^2 \left( \frac{1}{f_A^* p_A^* q_A^*} + \frac{1}{f_B^* p_B^* q_B} \right).
\]

Under the assumption of Theorem 2.1, we have the following results.

(a) For the allocation proportion \( N_{n,A}/n \), we have

\[
N_{n,A}/n - v_A^* = O\left( \frac{\log \log n}{n} \right) \text{ a.s. and } \sqrt{n}(N_{n,A}/n - v_A^*) \xrightarrow{D} N(0, \sigma_A^*)^2,
\]

where \( \sigma_A^* = f_A^*(1 - f_A^*) + 2\sigma_{MLE}^2 \).

(b) Given a covariate \( y \) with \( P(\eta = y) > 0 \), we have

\[
N_{n,A|y}/N_n(y) \rightarrow v_{A|y}^* \text{ a.s. and } \sqrt{n}(N_{n,A|y}/N_n(y) - v_{A|y}^*) \xrightarrow{D} N(0, \sigma_{A|y}^*)^2,
\]

where \( v_{A|y}^* = v_A^* \).

\[
(\sigma_{A|y}^*)^2 = f_A^*(1 - f_A^*) + 2\sigma_{MLE}^2 P(\eta = y)
- 2 f_A^* \left( (p_{A|y} - p_A^*) \frac{\partial f_A^*}{\partial p_A^*} + (p_{B|y} - p_B^*) \frac{\partial f_A^*}{\partial p_B^*} \right) P(\eta = y)
\]

and \( p_{k|y} = P(\text{success} | \eta = y \text{ and treatment } k), k = A, B; \)

(c) For the success proportion \( S_n/n \), we have

\[
S_n/n \rightarrow \mu_S^* \text{ a.s. and } \sqrt{n}(S_n/n - \mu_S^*) \xrightarrow{D} N(0, \sigma_S^*)^2,
\]

where \( \mu_S^* = f_A^* p_A^* + f_B^* p_B^* \) and

\[
\sigma_S^2 = \mu_S^*(1 - \mu_S^*) + 2(p_A^* - p_B^*)^2 \sigma_{MLE}^2 + 2(p_A^* - p_B^*) \left( p_A^* q_A^* \frac{\partial f_A^*}{\partial p_A^*} + p_B^* q_B^* \frac{\partial f_A^*}{\partial p_B^*} \right)
= \mu_S^*(1 - \mu_S^*) + 2(p_A^* - p_B^*)^2 \sigma_{MLE}^2.
\]

To compare the covariate-adjusted allocation rule (2.2) with the non covariate-adjusted rule (2.3), we consider the following hypothetical example.
**Example 1.** Assume that hypotensive patients arrive sequentially to be treated by two different drugs ($A$ and $B$). These patients are classified into two possible groups, with cerebrovascular disease (CVD) and non CVD (NCVD). For illustration, assume that CVD/NCVD is an important covariate and treatment efficacy may possibly behaves very different for these two groups. Hence, for the treatment assignment scheme, the covariate is dichotomous (CVD Vs NCVD). Let the success probabilities of treatments $A$ and $B$ for a CVD patient be $p_{A|CVD}$ and $p_{B|CVD}$ respectively, and the success probabilities of treatment $A$ and $B$ for a NCVD patient are $p_{A|NCVD}$ and $p_{B|NCVD}$ respectively. We choose the covariate vector $\eta$ to be $(1, 0)$ if a patient is CVD, and $(1, 1)$ if the patient is NCVD. Let $p = P(\text{CVD} | \text{patient})$ be the probability of a patient being CVD. We use two randomized procedures to randomize patients. One is (2.2) (referred to C) in which the covariate information is used. Another is (2.3) (referred to NC) in which the covariate information is not being used.

Table 1 gives the asymptotic means and related variances of success proportions $S_n/n$ and allocation proportions $N_{n,A}/n$ of these two procedures for different success probabilities and $p$. The success proportion in limit for the allocation rule C is always larger than the corresponding quantity for the rule NC if responses have treatment-covariate interactions, while, the allocation proportions are not significantly different for these two allocation rules.

Table 2 gives the asymptotic means and related variances of allocation proportions $N_{n,A|CVD}/N_n(CVD)$ (given the covariate CVD) and $N_{n,A|NCVD}/N_n(NCVD)$ (given the covariate NCVD) of this two allocation rules for different success probabilities. The proportions $N_{n,A|CVD}/N_n(CVD)$ and $N_{n,A|NCVD}/N_n(NCVD)$ are different and shift to a better treatment for the allocation rule $C$ when responses have treatment-covariate interactions, while, they are the same for the rule $NC$.

3. **General covariate-adjusted adaptive design.** A more general clinical scenario where more than two treatments are considered in this section. Given a clinical trial with $K$ treatments, suppose a patient with a covariate vector $\eta$ is assigned to treatment $k$, $k = 1, \ldots, K$, the observed response is $\xi_k$. Assume that the responses
Table 1
Asymptotic means and variances of success proportions and allocation proportions for (i) assignments using the covariate information (C), and (ii) assignments not using the covariate information (NC). Here $p = P(CVD \mid \text{patient})$.

| $(p_A|\text{CVD}, p_B|\text{CVD})$ | $p$ | $(\mu_S, \sigma^2)$ | $(\mu_S, \sigma^2)$ | $(v_A, \sigma^2)$ | $(v_A, \sigma^2)$ |
|-----------------|-----|-------------------|-------------------|-----------------|-----------------|
| $(p_A|\text{NCVD}, p_B|\text{NCVD})$ |     | $C$               | $NC$              | $C$             | $NC$            |
| (.9, .6)        | .5  | (.8013, 2536)     | (.7977, 2635)     | (.7922, 1.7354) | (.7907, 1.7993) |
| (.8, .6)        | .7  | (.8236, 2474)     | (.8206, 2568)     | (.8182, 1.5966) | (.8169, 1.6521) |
|                 | .3  | (.7790, 2589)     | (.7759, 2662)     | (.7662, 1.8729) | (.7650, 1.9248) |
| (.9, .7)        | .5  | (.8021, 2360)     | (.8006, 2377)     | (.7607, 2.1134) | (.7532, 2.1367) |
| (.8, .6)        | .7  | (.8248, 2226)     | (.8234, 2244)     | (.7741, 2.1275) | (.7672, 2.1531) |
|                 | .3  | (.7795, 2483)     | (.7783, 2494)     | (.7473, 2.0990) | (.7414, 2.1139) |
| (.9, .7)        | .5  | (.8021, 2360)     | (.7500, 1.875)    | (.5334, 2.1803) | (.5000, 2.9167) |
| (.6, .8)        | .7  | (.8248, 2226)     | (.7790, 1.893)    | (.6377, 2.1837) | (.6119, 2.9097) |
|                 | .3  | (.7795, 2483)     | (.7381, 2.089)    | (.4291, 2.1551) | (.3993, 2.6688) |
| (.6, .4) (.6, .4) | .5, .7, .3 | (.5385, 3195)     | (.5385, 3195)     | (.6923, 1.9882) | (.6923, 1.9882) |
| (.6, .4)        | .5  | (.5385, 3195)     | (.5000, 2.500)    | (.5000, 2.0251) | (.5000, 2.2500) |
| (.4, .6)        | .7  | (.5385, 3195)     | (.5064, 2625)     | (.5769, 2.0192) | (.5795, 2.2057) |
|                 | .3  | (.5385, 3195)     | (.5064, 2625)     | (.4231, 2.0192) | (.4205, 2.2057) |
| (.5, .5) (.5, .5) | .5, .7, .3 | (.5000, 2500)     | (.5000, 2500)     | (.5000, 2.2500) | (.5000, 2.2500) |
| (.4, .2)        | .5  | (.3021, 3200)     | (.2500, 1.875)    | (.4666, 2.9787) | (.5000, 2.9167) |
| (.1, .3)        | .7  | (.3195, 3190)     | (.2781, 2169)     | (.5709, 2.7845) | (.6007, 2.7626) |
|                 | .3  | (.2848, 3205)     | (.2390, 1.999)    | (.3623, 3.1513) | (.3881, 3.0611) |
Table 2
Asymptotic means and variances of allocation proportions for (i) assignments using the covariate information (C), and (ii) assignments not using the covariate information (NC), with covariate CVD or NCVD given and \( P(CVD|patient) = 1/2 \).

| \( (p_A|CVD, p_B|CVD) \) | \( (v_A|CVD, \sigma^2) \) | \( (v_A|NCVD, \sigma^2) \) | \( (v_A|NCVD, \sigma^2) \) |
|-----------------|-----------------|-----------------|-----------------|
| (.9, .6, .8, .6) | (.8571, 1.3858)  | (.7907, .9418)  | (.7273, 2.0766)  |
| (.9, .7, .8, .6) | (.7941, 2.1480)  | (.7532, 1.1275) | (.7273, 2.0766)  |
| (.9, .7, .6, .8) | (.7941, 2.1480)  | (.5000, 1.5167) | (.7273, 2.0766)  |
| (.6, .4, .6, .4) | (.6923, 1.9882)  | (.6923, 1.1006) | (.6923, 1.9882)  |
| (.6, .4, .6, .6) | (.6923, 1.9882)  | (.5000, 1.2000) | (.3077, 1.9882)  |
| (.5, .5, .5, .5) | (.5000, 2.2500)  | (.5000, 1.2500) | (.5000, 2.2500)  |
| (.4, .2, .1, .3) | (.7273, 2.4523)  | (.5000, 1.5167) | (.2059, 3.3693)  |

and the covariate vector satisfy the following equations as

\[
E[\xi_k|\eta] = p_k(\theta_k, \eta), \quad \theta_k \in \Theta_k, \quad k = 1, \ldots, K,
\]

where \( p_k(\cdot, \cdot), k = 1, \ldots, K, \) are known functions, \( \theta_k, k = 1, \ldots, K, \) are parameters whose values are unknown, \( \Theta_k \subset \mathbb{R}^d \) is the parameter space of \( \theta_k, k = 1, \ldots, K. \) Write \( \theta = (\theta_1, \ldots, \theta_K) \) and \( \Theta = \Theta_1 \times \cdots \times \Theta_K. \)

**Remark 3.1** This model includes the generalized linear models in McCullagh and Nelder (1989) as special cases. In Sections 4 and 5, two special cases, linear model and logit model, are considered.

### 3.1. Allocation rule and the design.

Suppose \( \{\xi_{m,k}, k = 1, \ldots, K, m = 1, 2, \ldots\} \) denote the responses, \( \{\eta_m, m = 1, 2, \ldots\} \) denote the covariates, where \( \xi_{m,k} \) is the response of the \( m \)-th subject on the treatment \( k, k = 1, \ldots, K, \) and \( \eta_m \) is the covariate of the \( m \)-th subject. In practical situations, only \( \xi_{m,k} \) with \( X_{m,k} = 1 \) is observed. We assume that \( \{\xi_{m,1}, \ldots, \xi_{m,K}, \eta_m\}, m = 1, 2, \ldots\) is a sequence of i.i.d. random vectors whose distributions are the same as that of \( (\xi_1, \ldots, \xi_K, \eta). \) And denote \( \xi_m = (\xi_{m,1}, \ldots, \xi_{m,K}). \) A general covariate-adjusted adaptive design is a design in which the information of the parameter \( \theta \) and the covariate information are used to randomize the assignment of treatments to the next patient. It is defined as follows.
Covariate-adjusted adaptive design. Assume that \( m \) \((m = 0, 1, \ldots)\) subjects are assigned and their responses \( \{\xi_j, j = 1, \ldots, m\} \) and the corresponding covariates \( \{\eta_j, j = 1, \ldots, m\} \) are observed. We let \( \hat{\theta}_m = (\hat{\theta}_{m,1}, \ldots, \hat{\theta}_{m,K}) \) be an estimate of \( \theta = (\theta_1, \ldots, \theta_K) \). Here for each \( k = 1, \ldots, K \), \( \hat{\theta}_{m,k} = \hat{\theta}_{0,k} = \theta_{0,k} \), \( k = 1, \ldots, K \), where \( \theta_0 = (\theta_{0,1}, \ldots, \theta_{0,K}) \) is given. Now, when the \((m + 1)\)-th subject is ready for randomization and its covariate \( \eta_{m+1} \) is recorded, we assign the patient to treatment \( k \) with a probability \( P_k = f_k(\hat{\theta}_m, \eta_{m+1}), k = 1, \ldots, K \). So,

\[
P_k = P(X_{m+1,k} = 1|F_m, \eta_{m+1}) = f_k(\hat{\theta}_m, \eta_{m+1}) \quad k = 1, \ldots, K,
\]

where \( F_m = \sigma(X_1, \ldots, X_m, \xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_m) \) is the sigma field of the history, \( f_k(\cdot, \cdot), k = 1, \ldots, K \) are some given functions. Given \( F_m \) and \( \eta_{m+1} \), the response \( \xi_{m+1} \) of the \((m + 1)\)-th subject is assumed to be independent of its assignment \( X_{m+1} \). We call the function \( f(\cdot, \cdot) = (f_1(\cdot, \cdot), \ldots, f_K(\cdot, \cdot)) \) the allocation function and it satisfies \( f_1 + \cdots + f_K \equiv 1 \).

Let \( g(\theta^*) = E[f(\theta^*, \eta)] \). Then

\[
P(X_{m+1,k} = 1|F_m) = g_k(\hat{\theta}_m), \quad k = 1, \ldots, K.
\]

Before deriving the asymptotic properties of the covariate-adjusted adaptive designs, we need to consider the selection of an appropriate allocation function.

3.2. The choice of allocation functions. Usually, the allocation function \( f = (f_1, f_2, \ldots, f_K) \) can be chosen as \( f_k(\theta, \eta) = R_k(\theta, \eta', \ldots, \theta_K \eta') \), \( k = 1, \ldots, K \), where \( R_k(\cdot), k = 1, \ldots, K \), are real functions defined on \( \mathbb{R}^K \) with

\[
0 < R_k(\cdot) < 1, \quad \sum_{k=1}^{K} R_k(\cdot) = 1 \quad \text{and} \quad R_i(x_1, \ldots, x_K) = R_j(x_1, \ldots, x_K) \quad \text{whenever} \quad x_i = x_j.
\]

Here, it is assumed that \( \eta \) and \( \theta_k, k = 1, \ldots, K \), have the same dimension. Otherwise, some modifications are necessary. We also assume that \( R_k(\cdot), k = 1, \ldots, K \), satisfy
two regular conditions: (i) They are continuous functions; (ii) For some $0 < \delta < 1$,
\[
\left| \frac{\partial R_k}{\partial x_i} \right|_x - \left| \frac{\partial R_k}{\partial x_i} \right|_y \leq C_0 \|x - y\|^\delta, \forall x \text{ and } y, \quad i, k = 1, \ldots, K.
\]
These two regular conditions, together, with $\mathbb{E}\|\eta\|^{2+\delta} < \infty$, implies Condition A (Given in Section 3.3) which is needed for deriving asymptotic properties, and
\[
\frac{\partial q_k}{\partial \theta^*} \bigg|_{\theta^* = \theta} = \mathbb{E} \left[ \frac{\partial R_k}{\partial x} \bigg|_{x = \theta, \eta'_k = 1, \ldots, K} \right].
\]

Given the previous responses and the covariate of the next subject for randomization, the assigning probabilities stated in (3.1) are:
\[
(3.4) \quad P(X_{m+1,k} = 1|\mathcal{F}_m, \eta_{m+1}) = R_k(\hat{\theta}_{m,1}\eta'_{m+1}, \ldots, \hat{\theta}_{m,K}\eta'_{m+1}), \quad k = 1, \ldots, K.
\]
In practice, the functions $R_k$ can be defined to be
\[
R_k(x) = \frac{G(x_k)}{G(x_1) + \cdots + G(x_K)}, \quad k = 1, \ldots, K,
\]
where $G$ is a real function defined on $\mathbb{R}$ satisfying $0 < G(x) < \infty$. For example we can define $R_k(x) = e^{Tx_k}/(e^{Tx_1} + \cdots + e^{Tx_K})$, $k = 1, \ldots, K$.

In the two-treatment case, one can let $R_1(x_1, x_2) = G(x_1 - x_2)$ and $R_2(x_1, x_2) = G(x_2 - x_1)$, where $G$ is real function defined on $\mathbb{R}$ satisfying $G(0) = 1/2$, $G(-x) = 1 - G(x)$ and $0 < G(x) < 1$ for all $x$. For the normal linear regression model, Bandyopadhyay and Biswas (2001) suggested to choose $G(x) = \Phi(x/T)$, where $\Phi(\cdot)$ is the standard normal distribution. For the logistic regression model, Rosenberger, Vidyashankar and Agarwal (2001) suggest using the estimated covariate-adjusted odds ratio to allocation subjects, which is equivalent to defining $R_k(x_1, x_2) = e^{x_k}/(e^{x_1} + e^{x_2})$, $k = 1, 2$ (See (2.2)).

3.3. Asymptotic properties. Write $f(\theta^*, y) = (f_1(\theta^*, y), \ldots, f_K(\theta^*, y))$, $g(\theta^* = (g_1(\theta^*), \ldots, g_K(\theta^*))$, $v_k = g_k(\theta) = \mathbb{E}[f_k(\theta, \eta)]$, $k = 1, \ldots, K$, and $v = (v_1, \ldots, v_K)$. We assume that $0 < v_k < 1$, $k = 1, \ldots, K$. For the allocation function $f(\theta^*, y)$ we assume the following condition.

**Condition A** We assume that the parameter spaces $\Theta_k$, $k = 1, \ldots, K$, are close sets in $\mathbb{R}^d$. 

14
(A1) For each fixed $y$, $f_k(\theta^*, y) > 0$ is a continuous function of $\theta^*$, $k = 1, \ldots, K$.

(A2) For each $k = 1, \ldots, K$, $f_k(\theta^*, \eta)$ is differentiable with respect to $\theta^*$ under the expectation, and there is a $\delta > 0$ such that

$$g_k(\theta^*) = g_k(\theta) + (\theta^* - \theta)\left(\frac{\partial g_k}{\partial \theta^*}\right)' + o(\|\theta^* - \theta\|^{1+\delta}),$$

where $\frac{\partial g_k}{\partial \theta^*} = (\partial g/\partial \theta^*_{11}, \ldots, \partial g/\partial \theta^*_{Kd})$.

**Theorem 3.1** Suppose

(3.5) \[ \hat{\theta}_n \to \theta \quad \text{a.s.} \]

Under Condition (A1), we have

\[ \frac{N_{n,k}}{n} \to v_k \quad \text{a.s.} \]

and

(3.6) \[ P(X_{n,k} = 1|F_{n-1}, \eta_n = y) \to f_k(\theta, y) \quad \text{a.s.,} \quad k = 1, \ldots, K. \]

Further, if

(3.7) \[ \hat{\theta}_n - \theta = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.,} \]

then under Conditions (A1)-(A2) we have

(3.8) \[ \frac{N_n}{n} - v = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.} \]

**Theorem 3.2** Suppose

(3.9) \[ \hat{\theta}_{nk} - \theta_k = \frac{1}{n} \sum_{m=1}^{n} X_{m,k}h_k(\xi_{m,k}, \eta_m) + o(n^{-1/2}) \quad \text{a.s.,} \quad k = 1, \ldots, K, \]

where $h_k$, $k = 1, \ldots, K$, are $K$ functions with $E[h_k(\xi_k, \eta)|\eta] = 0$, $k = 1, \ldots, K$. Also, we assume that $E[h_k(\xi_k, \eta)]^2 < \infty$ for some $\delta > 0$, $k = 1, \ldots, K$.

Under Condition A, we have

\[ \frac{N_n}{n} - v = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s. and} \quad \hat{\theta}_n - \theta = O\left(\sqrt{\frac{\log \log n}{n}}\right). \]
Further, let

$$V_k = E\{f_k(\theta, \eta)(h_k(\xi_k, \eta))'h_k(\xi_k, \eta)\}, \quad k = 1, \ldots, K,$$

$$V = \text{diag}(V_1, \ldots, V_K),$$

$$\Sigma_1 = \text{diag}(v) - v'v, \quad \Sigma_2 = \sum_{k=1}^K \frac{\partial g}{\partial \theta_k} V_k(\frac{\partial g}{\partial \theta_k})',$$

and $$\Sigma = \Sigma_1 + 2\Sigma_2.$$ Then

$$\sqrt{n}(N_n/n - v) \overset{D}{\to} N(0, \Sigma) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \overset{D}{\to} N(0, V).$$

Remark 3.2 Conditions (3.5), (3.7) and (3.9) depend on different estimation methods. In the next two sections, we will show that they are satisfied in many popular cases.

Remark 3.3 The following fact is worth noting. \{$(\xi_{m,k}, \eta_m) : \text{for which } X_{m,k} = 1, m = 1, \ldots, n, k = 1, \ldots, K$, are all sequences of neither independent nor identically distributed random vectors, regardless of whether $\{X_m\}$ is given. Hence, the estimators in adaptive designs do not directly inherit the properties of the related traditional estimators used in fixed assignment trials. Sometimes, their properties are quite different. See Remarks 4.2 and 5.1.

Theorems 3.1 and 3.2 provide us general results on the asymptotic properties of the proportions $N_{n,k}/n, k = 1, \ldots, K$. However, one may be more interested in the proportions when the covariate is given. Given a covariate $y$, the proportion of subjects assigned to treatment $k$ is

$$\frac{\sum_{m=1}^n X_{m,k} I\{\eta_m = y\}}{\sum_{m=1}^n I\{\eta_m = y\}} := \frac{N_{n,k|y}}{N_n(y)},$$

where $N_{n,k|y}$ is the number of subjects having covariate $y$ which are randomized to treatment $k, k = 1, \ldots, K$, in the $n$ trials, and $N_n(y)$ is the total number of subjects having covariate $y$. Write $N_{n|y} = (N_{n,1|y}, \ldots, N_{n,K|y})$. The following theorem establish the asymptotic results on these proportions.
Theorem 3.3 Given a covariate \( y \), suppose \( P(\eta = y) > 0 \). Under Conditions (A1) and (3.5),

\[
N_n, k|y / N_n(y) \to f_k(\theta, y) \quad \text{a.s.} \quad k = 1, \ldots, K.
\]

Under Conditions (A1)-(A2) and (3.9),

\[
\sqrt{N_n(y)} \left( N_n|y / N_n(y) - f(\theta, y) \right) \overset{D}{\to} N(0, \Sigma|y),
\]

where

\[
\Sigma|y = \text{diag}(f(\theta, y)) - f(\theta, y)'f(\theta, y) + 2 \sum_{k=1}^{K} \frac{\partial f(\theta, y)}{\partial \theta_k} V_k \left( \frac{\partial f(\theta, y)}{\partial \theta_k} \right) P(\eta = y).
\]

4. Linear Regression Model. Suppose the response \( \xi_k \) of a subject on the treatment \( k, k = 1, \ldots, K \), and its covariate \( \eta \) satisfies the linear regression model as

\[
E[\xi_k|\eta] = p_k(\theta_k, \eta) = \theta_k \eta', \quad k = 1, \ldots, K,
\]

where \( \theta_k = (\theta_{k1}, \ldots, \theta_{kd}), k = 1, \ldots, K \), are regression coefficients. Let \( \hat{\theta}_{nk} \) minimize the error sum of squares

\[
S_k(\theta) = \sum_{m=1}^{n} X_{m,k}(\xi_m - \theta_k \eta'_m)^2 \quad \text{over} \quad \theta_k \in \Theta_k,
\]

\( k = 1, \ldots, K \). Here \( \hat{\theta}_{nk} \) is the least square estimator.

We have the following asymptotic properties.

Theorem 4.1 Suppose the following conditions are satisfied:

(a) \( E[\|\eta\|^{2+\epsilon}] < \infty, E[\|\xi_{m,k}\eta\|^{2+\epsilon}] < \infty \) for some \( 0 < \epsilon \leq 1, k = 1, \ldots K \);

(b) The matrix \( E[\eta'\eta] \) is positive definite;

(c) For each \( k = 1, \ldots, K \), the parameter space \( \Theta_k \) is a bounded close domain in \( \mathbb{R}^d \), and the true value \( \theta_k \) is an interior point of \( \Theta_k \);
Under Condition (A1), we have for $k = 1, \ldots, K$,

\begin{equation}
\frac{N_{n,k}}{n} \to v_k \text{ a.s.}, \quad P(X_{n,k} = 1) \to v_k
\end{equation}

and

\[ P[X_{n,k} = 1 | \mathcal{F}_{n-1}, \eta_n = y] \to f_k(\theta, y) \text{ a.s.} \]

Under Conditions (A1)-(A2), we have

\begin{equation}
\frac{N_n}{n} - v = O\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ a.s.}, \quad \hat{\theta}_n - \theta = O\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ a.s.}
\end{equation}

and

\begin{equation}
\sqrt{n}(N_n/n - v) \overset{D}{=} N(0, \Sigma) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \overset{D}{=} N(0, V),
\end{equation}

where $\Sigma$ and $V$ are defined in Theorem 3.2 with $V_k = \Gamma_{\eta k}^{-1} \Gamma_{\xi k}^{-1}$,

\[ I_{\eta k} = \mathbb{E}[f_k(\theta, \eta)\eta'] \quad \text{and} \quad I_{\xi k} = \mathbb{E}[f_k(\theta, \eta)(\xi_k - \theta' \eta)^2 \eta'], \]

$k = 1, \ldots, K$. Moreover, if $P(\eta = y) > 0$ for given $y$, then (3.10) and (3.11) hold.

**Remark 4.1** Theorem 1 of Bandyopadhyay and Biswas (2001) is a special case of the results of our Theorem 4.1, in which $\theta_{11} = \mu_1$, $\theta_{21} = \mu_2$, $\theta_{1j} = \theta_{2j}$, $j = 2, \ldots, d$, and the first component of $\eta$ is 1. Bandyopadhyay and Biswas (2001) only studied the consistency of $N_{n,1}/n$ and $P(X_{n,1} = 1)$, and their applications are limited to the normal linear model.

**Remark 4.2** From Theorem 4.1, it follows that

\begin{equation}
\sqrt{N_{n,k}} (\hat{\theta}_{n,k} - \theta_k) \overset{D}{=} N(0, v_k V_k), \quad k = 1, \ldots, K.
\end{equation}

It should be note that the asymptotic variances are different from the ordinary linear models. For the latter, we have

\[ \sqrt{N_{n,k}} (\hat{\theta}_{n,k} - \theta_k) \overset{D}{=} N(0, [\mathbb{E}\eta']^{-1} \text{Var}((\xi_k - \theta_k \eta)\eta) [\mathbb{E}\eta']^{-1}), \quad k = 1, \ldots, K. \]
If the allocation functions \( f_k(\theta, \eta) \) do not depend on \( \eta \), then \( f_k(\theta, \eta) = g_k(\theta) = v_k \), and so (4.5) and (4.6) are the same. So, if the assignments do not depend on the covariate, the adaptive procedure does not influence the asymptotic behaviors of the estimators. While, if the assignments depend on the covariate, the adaptive procedure influences the asymptotic variances of the estimators. But in both cases, the exact distributions of the estimators in the adaptive design are different from those of traditional estimators.

**Remark 4.3** The value of \( V \) is important for deriving the asymptotic variance. An estimator of \( V \) can be defined as the following procedure. First, we estimate \( \hat{I}_{n,\eta k} \) by

\[
\hat{I}_{n,\eta k} = \frac{1}{n} \sum_{m=1}^{n} X_{m,k} \eta_m \eta_m.
\]

Then, we estimate \( \hat{I}_{n,\xi k} \) by

\[
\hat{I}_{n,\xi k} = \frac{1}{n} \sum_{m=1}^{n} X_{m,k} (\xi_{m,k} - \hat{\theta}_{n,k} \eta_m')^2 \eta_m \eta_m.
\]

\( V \) can be estimated by

\[
\hat{V}_n = \text{diag} (\hat{I}_{n,\eta 1}^{-1}, \hat{I}_{n,\xi 1}^{-1}, \ldots, \hat{I}_{n,\eta K}^{-1}, \hat{I}_{n,\xi K}^{-1}).
\]

With the estimator of \( V \), one can easily get the estimates of the asymptotic variance of the proportions \( N_{n,k}/n \) and the proportions for given covariates.

5. Binary Responses with Logistic Regression Covariate. We now consider the case of dichotomous (i.e. success or failure) responses. Let \( \xi_k = 1 \) if a subject on treatment \( k \) is a success and 0 otherwise, \( k = 1, \ldots, K \). Assume that the responses and the covariate satisfy the logistic regression model

\[
P(\xi_k = 1|\eta) = p_k(\theta_k, \eta) = \frac{\exp\{\theta_k \eta'\}}{1 + \exp\{\theta_k \eta'\}}, \quad k = 1, \ldots, K.
\]

Denote \( p_k = p_k(\theta_k, \eta), q_k = 1 - p_k, k = 1, \ldots, K \). By the method of maximum likelihood, one is able to estimate the parameters \( \theta_k, k = 1, \ldots, K \). For each \( k = 1, \ldots, K \), let \( p_{m,k} = p_k(\theta_k, \eta_k) \) and

\[
L_k = \prod_{m=1}^{n} p_{m,k}^X_{m,k} (1 - p_{m,k})^{X_{m,k}(1 - \xi_{m,k})}.
\]
The maximum likelihood estimator \( \hat{\theta}_{n,k} \) of \( \theta_k \) \((k = 1, \ldots, K)\) is that for which

(5.1) \[ \hat{\theta}_{n,k} \text{ maximizes } L_k \text{ over } \theta_k \in \Theta_k. \]

**Theorem 5.1** Suppose Condition A and the following conditions are satisfied:

(i) \( E\|\eta\|^2 + \epsilon < \infty \) for some \( \epsilon > 0 \); The matrix \( E[\eta'\eta] \) is positive definite;

(ii) For each \( k = 1, \ldots, K \), the parameter space \( \Theta_k \) is a bounded close domain in \( \mathbb{R}^d \), and the true values \( \theta_k \) is an interior point of \( \Theta_k \).

Let \( v_k = E[f_k(\theta, \eta)] \), \( k = 1, \ldots, K \). Define

\[ I_k = I_k(\theta) = E\{f_k(\theta, \eta)p_kq_k\eta'\eta\}, \quad k = 1, \ldots, K. \]

Then
\[ P(X_{n,k} = 1|\mathcal{F}_{n-1}, \eta_n = y) \to f_k(\theta, y) \quad \text{a.s.,} \quad k = 1, \ldots, K \]

and

(5.2) \[ \frac{N_n}{n} - v = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.} \quad \text{and} \quad \hat{\theta}_n - \theta = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.} \]

Also,
\[ \sqrt{n}(N_n/n - v) \overset{D}{\to} N(0, \Sigma) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta) \overset{D}{\to} N(0, V), \]

where \( V \) and \( \Sigma \) are defined in Theorem 3.2 with \( V_k = I_k^{-1} \), \( k = 1, \ldots, K \).

Moreover, if \( P(\eta = y) > 0 \) for given covariate \( y \), then (3.10) and (3.11) hold.

An estimator of \( I_k \) is
\[ \hat{I}_{n,k} = \frac{1}{n} \sum_{m=1}^{n} X_{m,k} p_k(\hat{\theta}_{n,k}, \eta_m)(1 - p_k(\hat{\theta}_{n,k}, \eta_m))\eta_m'\eta_m, \]
\( k = 1, \ldots, K \). And then an estimator of \( V \) is
\[ \hat{V}_n = \text{diag}(\hat{I}_{n,1}^{-1}, \ldots, \hat{I}_{n,K}^{-1}). \]
Remark 5.1 From Theorem 5.1, it follows that
\[ \sqrt{N_{n,k}} (\hat{\theta}_{n,k} - \theta_k) \xrightarrow{D} N(0, v_k I_k^{-1}) \quad k = 1, \ldots, K. \]

The above asymptotic normalities are different from related traditional ones as
\[ \sqrt{N_{n,k}} (\hat{\theta}_{n,k} - \theta_k) \xrightarrow{D} N(0, \{E[p_k q_k \eta' \eta] \}^{-1}), \quad k = 1, \ldots, K. \]

Also, when the assignments do not depend on the covariate \( \eta \), the two kind of asymptotic normalities are the same.

Remark 5.2 For the general non-linear regression model, under suitable regular conditions we have similar asymptotic properties.

The last theorem gives us the asymptotic normality of the total numbers of successes.

Theorem 5.2 Let \( S_n = \sum_{m=1}^{n} \sum_{k=1}^{K} X_{m,k} \xi_{m,k} \) be the total numbers of successes in \( n \) trials. Then under the conditions in Theorem 5.1,
\[ \frac{S_n}{n} - \mu_S = O\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ a.s. and } \sqrt{n}(S_n/n - \mu_S) \xrightarrow{D} N(0, \sigma_S^2), \]
where
\[ \mu_S = \sum_{k=1}^{K} E[f_k p_k] \]
and
\[ \sigma_S^2 = \mu_S(1 - \mu_S) + 2 \sum_{j,k=1}^{K} E[f_j p_j q_j \eta] \mathbf{I}_j^{-1} E[\left(\frac{\partial f_k}{\partial \theta_j}\right)' p_k] \]
\[ + 2 \sum_{j=1}^{K} \left[ \sum_{k=1}^{K} E[\frac{\partial f_k}{\partial \theta_j} p_k] \right] \mathbf{I}_j^{-1} \left[ \sum_{k=1}^{K} E[\left(\frac{\partial f_k}{\partial \theta_j}\right)' p_k] \right] \]
and \( f_k = f_k(\theta, \eta), \quad k = 1, \ldots, K. \)

6. Discussion. The main contributions of this paper are two-fold. First, a covariate-adjusted design is derived to serve as a paradigm for treatment allocation procedures in clinical studies where covariates are available. Second, asymptotic
properties are given to provide the statistical basis for inferences related to treatment efficacies. For the comparison of two treatments with binary responses, we demonstrate the two important properties of the covariate-adjusted adaptive design (targeting the odd ratio): (i) it has higher total number of successes than adaptive designs ignoring the information of the covariates; and (ii) it sends more subjects to the better treatment for a given covariate.

Treatment allocation problem in the presence of covariates have been considered by a number of researchers. For example, Zelen (1974) and Pocock and Simon (1975) considered balancing covariates by using idea of biased coin design (Efron, 1971). Atkinson (1982, 1999) tackled this problem by using $D$-optimality with a linear model. In addition, a one-arm bandit problem with covariate has also been applied by by Woodroofe (1979, 1982) and Sarkar (1991). However, all these efforts have not incorporated patients’ responses in the scheme of treatment allocation. In this paper, we have departed from early methodologies and considered treatment allocation schemes as a function of both patient responses and the covariates.

When covariate information is not being used in the treatment allocation scheme, optimal allocation proportion are usually determined with the assistance of some optimality criteria. Jennison and Turnbull (2000) described a general procedure to find optimal allocation. For covariate-adjusted adaptive design, it is still unclear how to define and how to search for an optimal allocation. The allocation function, $f_A(\theta, \eta)$ in Section 2, usually depends on the target allocation. For binary responses with two treatments, we use the odd ratio, which is also used in Rosenberger, Vidyashankar and Agarwal (2001). We may also use some other allocation proportions, for example, the three proportions considered in Hu and Rosenberger (2003).

For adaptive designs without covariates, Hu and Rosenberger (2003) studied the relationship among the power, the target allocation and the variability of the design. It is important to study the behavior of the power function, when a covariate-adjusted adaptive design is used in clinical trials. This will be a interesting topic for future research.
The proofs of the theorems are organized as following. First we prove the theorems for the general covariate-adjusted adaptive design in Section 3. Then we use these results to show theorems in Section 4 and 5. Finally, we apply the theorems in Section 5 to show the special cases (theorems in Section 2).

**Proof of Theorem 3.1.** It suffices to show \( N_n / n \to v \) a.s. under (3.5), and (3.8) under (3.7). Notice that for each \( k = 1, \ldots, K \),

\[
X_{m+1,k} = X_{m+1,k} - E[X_{m+1,k} | F_m] + g_k(\hat{\theta}_m)
\]

and then

\[
N_{n,k} = E[X_{1,k} | F_0] + \sum_{m=1}^{n} (X_{m,k} - E[X_{m,k} | F_{m-1}]) + \sum_{m=1}^{n} g_k(\hat{\theta}_m).
\]

Since \( \{X_{m,k} - E[X_{m,k} | F_{m-1}]\} \) is a sequence of bounded martingale differences. By the law of the iterated logarithm for martingales, it follows that

\[
\sum_{m=1}^{n} (X_{m,k} - E[X_{m,k} | F_{m-1}]) = O\left(\sqrt{\log \log n / n}\right) \text{ a.s.}
\]

And the results follow easily. \( \square \)

**Proof of Theorem 3.2.** Write \( \Delta_{M,m,k} = X_{m,k} - E[X_{m,k} | F_{m-1}] \), \( \Delta T_{m,k} = X_{m,k} h_k(\xi_k, \eta_m) \), \( k = 1, \ldots, K \). Let \( M_n = \sum_{m=1}^{n} \Delta M_m \) and \( T_n = \sum_{m=1}^{n} \Delta T_m \), where \( \Delta M_m = (\Delta M_{m,1}, \ldots, \Delta M_{m,K}) \) and \( \Delta T_m = (\Delta T_{m,1}, \ldots, \Delta T_{m,K}) \). Then \( \{(M_n, T_n)\} \) is a multi-dimensional martingale sequence satisfying the following conditions:

\[
|\Delta M_{n,k}| \leq 1 \quad \text{and} \quad E[\|\Delta T_{n,k}\|^{2+\delta}] \leq E[\|h_k(\xi_k, \eta)\|^{2+\delta}], \quad k = 1, \ldots, K;
\]

\[
E[(\Delta M_n)'\Delta M_n | F_{n-1}] = \text{diag}(g(\hat{\theta}_{n-1})) - (g(\hat{\theta}_{n-1}))'g(\hat{\theta}_{n-1}) \to \Sigma_1 \quad \text{in } L_1;
\]

\[
E[(\Delta T_{n,k})'\Delta T_{n,k} | F_{n-1}] = E_f[\hat{f}_k(\hat{\theta}_{n-1}, \eta) (h_k(\xi_k, \eta))' h_k(\xi_k, \eta)] \to V_k \quad \text{in } L_1, \quad k = 1, \ldots, K;
\]

\[
E[(\Delta M_{n,i})'\Delta T_{n,j} | F_{n-1}] = 0 \quad \text{for all } i, j \quad \text{and} \quad E[(\Delta T_{n,i})'\Delta T_{n,j} | F_{n-1}] = 0 \quad \text{for all } i \neq j.
\]

So, according to the law of the iterated logarithm for martingales, we have

\[
M_n = O\left(\sqrt{n \log \log n}\right) \text{ a.s. and} \quad T_n = O\left(\sqrt{n \log \log n}\right) \text{ a.s.};
\]
and, according to the weak convergence for martingales (c.f. Hall and Heyde (1980)), we have

\[(A.3) \quad n^{-1/2}(M_{[nt]}, T_{[nt], 1}, \ldots, T_{[nt], K}) \Rightarrow (W_t \Sigma_1^{1/2}, B_{t, 1} V_1^{1/2}, \ldots, B_{t, K} V_K^{1/2}),\]

where \(W_t\) is an \(K\)-dimensional Wiener process, \(B_{t, k}\), \(k = 1, \ldots, K\), are \(d\)-dimensional Wiener processes, and all these Wiener processes are independent. From (A.2), (A.3) and (3.9), it follows that

\[(A.4) \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, \Sigma) \quad \text{and} \quad \hat{\theta}_n - \theta = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.}\]

Next, we consider the proportion \(N_n/n\). From (A.1), (A.4) and the condition (A2), it follows that

\[N_{n, k} - n v_k = M_{n, k} + \sum_{m=1}^{n-1} \sum_{j=1}^{K} (\hat{\theta}_{m, j} - \theta_j) \left(\frac{\partial g_k}{\partial \theta_j}\right)' + \sum_{m=1}^{n-1} o(||\hat{\theta}_m - \theta||^{1+\delta})\]

That is

\[(A.5) \quad N_n - n v = M_n + \sum_{m=1}^{n} \sum_{j=1}^{K} \frac{1}{m} T_{m, j} \left(\frac{\partial g_k}{\partial \theta_j}\right)' + o(n^{1/2}) \quad \text{a.s.} \quad k = 1, \ldots, K.\]

Combing (A.2) and (A.5) yields

\[N_n - n v = O\left(\sqrt{n \log \log n}\right) \quad \text{a.s.}\]

Combing (A.3) and (A.5) yields

\[(A.6) \quad n^{-1/2}(N_{[nt]} - n t v) \Rightarrow W_t \Sigma_1^{1/2} + \int_0^t \sum_{j=1}^{K} \frac{B_{x, j}}{x} \, dx V_k^{1/2} \left(\frac{\partial g}{\partial \theta_j}\right)'\]

The co-variance matrix of the above Gaussian process is

\[t \Sigma_1 + \sum_{j=1}^{K} \int_0^t \int_0^t \frac{x \wedge y}{xy} \, dx dy \left(\frac{\partial g}{\partial \theta_j}\right) \Sigma_j \left(\frac{\partial g}{\partial \theta_j}\right)' = t(\Sigma_1 + 2 \Sigma_2) = t \Sigma.\]

The proof of Theorem 3.2 is now completed. \(\square\)
By computing the co-variance matrix of the above Gaussian process, we obtain

\begin{equation}
\frac{1}{n} \sum_{m=1}^{n} I\{\eta_m = y\} \rightarrow P(\eta = y) \text{ a.s.}
\end{equation}

and

\begin{align*}
\frac{1}{n} \sum_{m=1}^{n} X_{m,k} I\{\eta_m = y\} &= \frac{1}{n} \sum_{m=1}^{n} \left( X_{m,k} I\{\eta_m = y\} - E[X_{m,k} I\{\eta_m = y\} | \mathcal{F}_{m-1}] \right) + \frac{1}{n} \sum_{m=1}^{n} f_k(\theta_{m-1}, y) P(\eta_m = y) \\
&\rightarrow f_k(\theta, y) P(\eta = y) \text{ a.s.,}
\end{align*}

according the law of large numbers. (3.10) is proved. With the same argument as deriving (A.5), we can obtain

\begin{align*}
\zeta_{n,k}(y) := & \sum_{m=1}^{n} (X_{m,k} - f_k(\theta, y)) I\{\eta_m = y\} \\
= & \sum_{m=1}^{n} (\Delta \zeta_{n,k}(y) - E[\Delta \zeta_{n,k}(y) | \mathcal{F}_{n-1}]) + \sum_{m=1}^{n} (f_k(\theta_{m-1}, y) - f_k(\theta, y)) P(\eta = y) \\
= & \sum_{m=1}^{n} (\Delta \zeta_{m,k}(y) - E[\Delta \zeta_{m,k}(y) | \mathcal{F}_{m-1}]) + \sum_{j=1}^{K} \sum_{m=1}^{n} T_{m,j} \frac{\partial f_k(\theta, y)}{\partial \theta_j} P(\eta = y) + o(n^{1/2}) \text{ a.s.}
\end{align*}

Let \( \Delta \overline{M}_{n,k}(y) = \Delta \zeta_{n,k}(y) - E[\Delta \zeta_{n,k}(y) | \mathcal{F}_{n-1}] \), \( m = 1, 2, \ldots \). It is easily seen that

\begin{align*}
E[(\Delta \overline{M}_{n,k}(y))^2 | \mathcal{F}_{n-1}] &\rightarrow f_k(\theta, y) (1 - f_k(\theta, y)) P(\eta = y) = E[(W_{1,k}(y))^2], \ k = 1, \ldots, K, \\
E[\Delta \overline{M}_{n,k}(y) \Delta \overline{M}_{n,j}(y) | \mathcal{F}_{n-1}] &\rightarrow -f_k(\theta, y) f_j(\theta, y) P(\eta = y) = E[W_{1,k}(y)W_{1,j}(y)] \ \forall k \neq j
\end{align*}

and

\begin{equation}
E[\Delta \overline{M}_{n,k}(y) \Delta T_{n,j} | \mathcal{F}_{n-1}] = 0 = E[W_{1,k}(y)B_{t,j} V_j^{1/2}], \forall i, j = 1, \ldots, K,
\end{equation}

all in \( L_1 \), where \( B_{t,j}, \ j = 1, \ldots, K \) are dependent \( d \)-dimensional standard Wiener processes defined as in (A.3), and \( W_t = (W_{t,1}, \ldots, W_{t,K}) \) is a \( K \) dimensional Wiener processes which is independent of \( B_{t,j}, \ j = 1, \ldots, K \). Following the same argument as deriving (A.6), one can obtain

\begin{equation}
n^{-1/2}(\zeta_{nt,1}(y), \ldots, \zeta_{nt,K}(y)) \Rightarrow W_t + \sum_{j=1}^{K} \int_0^t \frac{B_{x,j} V_j^{1/2}}{x} dx \left( \frac{\partial f(\theta, y)}{\partial \theta_j} \right) P(\eta = y).
\end{equation}

By computing the co-variance matrix of the above Gaussian process, we obtain

\begin{equation}
n^{-1/2}(\zeta_{n,1}(y), \ldots, \zeta_{n,K}(y)) \overset{D}{\rightarrow} N(0, \Sigma_\eta P(\eta = y)).
\end{equation}
Notice
\[ \sqrt{N_n(y)} \left( \frac{N_n(y)}{N_n(y)} - f(\theta, y) \right) = \sqrt{\frac{n}{N_n(y)}} \frac{\zeta_{n,k}(y)}{\sqrt{n}}, \quad k = 1, \ldots, K \]
and (A.7). (3.11) is now proved. □

**Proof of Theorem 4.1.** By Theorems 3.1 and 3.2, it is sufficient to verify conditions (3.5) and (3.9). For each \( k = 1, \ldots, K \), let \( \hat{\theta}_{n,k}^* \) minimize the error sum of squares
\[ S_k(\theta_k) = \sum_{m=1}^n X_{m,k}(\xi_{m,k} - \theta_k \eta'_m)^2 \text{ over } \theta_k \in \mathbb{R}^d. \]
Then \( \hat{\theta}_{n,k}^* \) is the solution of the normal equation as
\[ \hat{\theta}_{n,k}^* \sum_{m=1}^n X_{m,k} \eta'_m \eta_m = \sum_{m=1}^n X_{m,k} \xi_{m,k} \eta_m. \]
Next, we show that
\[ (A.8) \quad \hat{\theta}_{n,k}^* \rightarrow \theta_k \quad \text{a.s.,} \quad k = 1, \ldots, K, \]
which implies \( \hat{\theta}_{n,k}^* \in \Theta_k, \quad k = 1, \ldots, K \), for \( n \) large enough. And so
\[ \hat{\theta}_{n,k} = \hat{\theta}_{n,k}^* \rightarrow \theta_k \quad \text{a.s.,} \quad k = 1, \ldots, K. \]
Notice that
\[ (A.9) \quad (\hat{\theta}_{n,k}^* - \theta_k) \frac{1}{n} \sum_{m=1}^n X_{m,k} \eta'_m \eta_m = \frac{1}{n} \sum_{m=1}^n X_{m,k}(\xi_{m,k} - \theta_k \eta'_m) \eta_m. \]
Since
\[ E[X_{m,k}(\xi_{m,k} - \theta_k \eta'_m) \eta_m | \mathcal{F}_{m-1}] = E[E[X_{m,k}(\xi_{m,k} - \theta_k \eta'_m) \eta_m | \mathcal{F}_{m-1}, \eta_m] | \mathcal{F}_{m-1}] = E\{f_k(\hat{\theta}_{m-1}, \eta_m) (E(\xi_{m,k} | \eta_m) - \theta_k \eta'_m) \eta_m | \mathcal{F}_{m-1} = 0, \]
\( \{X_{m,k}(\xi_{m,k} - \theta_k \eta'_m) \eta_m\} \) is a sequence of martingale differences with
\[ E\|X_{m,k}(\xi_{m,k} - \theta_k \eta'_m) \eta_m\|^{2+\delta} \leq 2E\|\xi_{m,k} \eta_m\|^{2+\delta} < \infty. \]
By the law of the iterated logarithm for martingales, we have
\[ (A.10) \quad \frac{1}{n} \sum_{m=1}^n X_{m,k}(\xi_{m,k} - \theta_k \eta'_m) \eta_m = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.} \]
Next we consider the term $\frac{1}{n} \sum_{m=1}^{n} X_{m,k} \eta'_m \eta_m$. It is obviously that $\{X_{m,k} \eta'_m \eta_m - \mathbb{E}[X_{m,k} \eta'_m \eta_m | \mathcal{F}_{m-1}]\}$ is also a sequence of martingale difference with $\mathbb{E}\|X_{m,k} \eta'_m \eta_m\|^{1+\delta/2} < \infty$. It follows from the law of large numbers for martingales that for some $\delta' > 0$,

$$
\frac{1}{n} \sum_{m=1}^{n} X_{m,k} \eta'_m \eta_m = \frac{1}{n} \sum_{m=1}^{n} \mathbb{E}[X_{m,k} \eta'_m \eta_m | \mathcal{F}_{m-1}] + o(n^{-\delta'})
$$

(A.11)

where $M_k(\theta^*) = \mathbb{E}[f_k(\theta^*, \eta) \eta' \eta]$. Notice that $\mathbb{E}[\eta' \eta]$ is positive definite, and $f_k(\theta^*, \eta) > 0$ for each $\eta$, and $\theta^*$. It follows that $M_k(\theta^*)$ is positive definite for each $\theta^*$. Further, it is easy seen that $M_k(\theta^*)$ is continuous. It follows that $M_k(\theta^*)$ is positive definite uniformly on bounded set $\Theta$. That is

$$
u M_k(\theta^*) u' \geq c_0 > 0 \text{ for all } \theta^* \in \Theta \text{ and } u \text{ with } \|u\| = 1.
$$

We conclude that $\frac{1}{n} \sum_{m=1}^{n} M_k(\hat{\theta}_{m-1})$ is positive definite uniformly in $n$ for $n$ large enough, which together with (A.11) implies that $\frac{1}{n} \sum_{m=1}^{n} X_{m,k} \eta'_m \eta_m$ is almost surely positive definite uniformly in $n$ for $n$ large enough. So

$$
\limsup_{n \to \infty} \left\| \frac{1}{n} \sum_{m=1}^{n} X_{m,k} \eta'_m \eta_m \right\|^{-1} < \infty \; \text{ a.s.}
$$

(A.12)

Combing (A.9), (A.10) and (A.12) yields

$$
\hat{\theta}_{n,k} - \theta_k = O\left(\sqrt{\frac{\log \log n}{n}}\right) \; \text{ a.s.,} \quad k = 1, \ldots, K.
$$

(A.8) is now proved, and we conclude that $\hat{\theta}_n = \hat{\theta}^*_n$ for $n$ large enough, and

$$
\hat{\theta}_n - \theta = O\left(\sqrt{\frac{\log \log n}{n}}\right) \; \text{ a.s.}
$$

(A.13)

By Theorem 3.1, (4.2) and (4.3) are proved.

Now, we prove (4.4). Notice (A.13). It is easily seen that for some $\delta' > 0$,

$$
M_k(\hat{\theta}_n) = M_k(\theta) + o(n^{-\delta'}) = I_{\eta,k} + o(n^{-\delta'}) \; \text{ a.s.,}
$$

which, together with (A.11), implies

$$
\frac{1}{n} \sum_{m=1}^{n} X_{m,k} \eta'_m \eta_m = I_{\eta,k} + o(n^{-\delta'}) \; \text{ a.s.}
$$

(A.14)
It follows from (A.9) that
\[ \hat{\theta}_{n,k} - \theta_k = \hat{\theta}^*_{n,k} - \theta_k = \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - \theta_k \eta'_m)\eta_m(\mathbf{I}_n^{-1} + o(n^{-\delta'})) \]
\[ = \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - \theta_k \eta'_m)\eta_m \mathbf{I}_{n,k}^{-1} + o(n^{-1/2}) \text{ a.s.} \]

(3.9) is satisfied with \( h_k = (\xi_{m,k} - \theta_k \eta'_m)\eta_m, \ k = 1, \ldots, K \). Now, (4.4) is proved by Theorem 3.2. □

**Proof of Theorem 5.1.** By Theorems 3.1 and 3.2, it suffices for us to verify (3.9). We will show that for some \( \tau > 0 \),

\[ \hat{\theta}_{n,k} - \theta_k = \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - p_k(\theta_k, \eta_m))\eta_m \mathbf{I}_{n,k}^{-1} + o(n^{-1/2}) \text{ a.s.} \]

for \( k = 1, \ldots, K \). We prove (A.15) step by step. In the first step, we show the consistency of \( \hat{\theta}_{n,k}, k = 1, \ldots, K \). In the second step, we show the condition (3.7). At the last step, we show the condition (A.15).

Now, we begin the first step. For each \( k = 1, \ldots, K \), let \( \hat{\theta}^*_{n,k} \) be such that

(\[ A.16 \] ) \[ \hat{\theta}^*_{n,k} \text{ maximizes } L_k \text{ over } \theta_k \in \mathbb{R}^d. \]

We show that

(\[ A.17 \] ) \[ \hat{\theta}^*_{n,k} \to \theta_k \text{ a.s.,} \]

which implies that, for \( n \) large enough, \( \hat{\theta}^*_{n,k} \in \Theta_k \), and then

(\[ A.18 \] ) \[ \hat{\theta}_{n,k} = \hat{\theta}^*_{n,k} \to \theta_k \text{ a.s.} \]

Obviously, \( \hat{\theta}^*_{n,k} \) is a solution of the equation as

(\[ A.19 \] ) \[ \frac{\partial}{\partial \theta_k} \left\{ \log \prod_{m=1}^{n} [p_k(\theta_k, \eta_m)]^{X_{m,k}\xi_{m,k}} [q_k(\theta_k, \eta_m)]^{X_{m,k}(1-\xi_{m,k})} \right\} = 0. \]

Let \( p_{m,k} = p_k(\theta_k, \eta_m), q_{m,k} = 1 - p_{m,k} \) and

\[ h(\theta_k) = \log \prod_{m=1}^{n} p_{m,k}^{X_{m,k}\xi_{m,k}} q_{m,k}^{X_{m,k}(1-\xi_{m,k})}. \]
Then
\[
\frac{\partial h}{\partial \theta_k} = \sum_{m=1}^n X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m,
\]
(A.20)
\[
\frac{\partial^2 h}{\partial \theta_k^2} = -\sum_{m=1}^n X_{m,k} p_{m,k} \eta_m' \eta_m.
\]

Obviously, \(\frac{\partial^2 h}{\partial \theta_k^2}\) is a non-negatively definite matrix. Notice
\[
P(X_{m,k}\eta_m' \eta_m \neq 0|F_{m-1}) = P(\eta_m \neq 0)P(X_{m,k} = 1|F_{m-1})
= P(\eta \neq 0)g_k(\hat{\theta}_{m-1}) \geq P(\eta \neq 0) \min_{\theta \in \Theta} g_k(\theta) \geq c_0P(\eta \neq 0) > 0.
\]

We have \(P(X_{m,k}\eta_m' \eta_m \neq 0, \text{i.o.}) = 1\) by the generalized Borel-Cantelli lemma (c.f., Corollary 2.3 of Hall and Heyde, 1980). We conclude that, with probability 1 for \(n\) large enough, \(\frac{\partial^2 h}{\partial \theta_k^2}\) is positive definite for every \(\theta_k\). It follows that equation (A.19) has a unique solution. Next, it is enough to prove that, with probability 1, in any neighborhood of the true value \(\theta_k\), there is a solution of equation (A.19) if \(n\) is large enough. It is obvious that
\[
\frac{\partial p_k}{\partial \theta_k} = p_k(1 - p_k)\eta, \quad \frac{\partial^2 p_k}{\partial \theta_k^2} = (1 - 2p_k)p_k(1 - p_k)\eta' \eta
\]

and
\[
|p_k(\theta^*, \eta) - p_k(\theta, \eta)| \leq \|\theta^* - \theta\|\delta\|\eta\|\delta \quad \forall 0 \leq \delta \leq 1.
\]

It follows that
\[
\frac{p_k(\theta^*, \eta)(1 - p_k(\theta^*, \eta)) - p_k(\theta, \eta)(1 - p_k(\theta, \eta))}{\|p_k(\theta^*, \eta) - p_k(\theta, \eta)\|} \leq 2\|\theta^* - \theta\|\delta\|\eta\|\delta.
\]

With the same argument for showing (A.12), one can show that
\[
\lim_{n \to \infty} \|\left[\frac{\partial^2 h}{n \partial \theta_k^2}\right]_{\theta_k}^{-1}\| < K_1 < \infty.
\]

On the other hand, by the Taylor theorem,
\[
\left.\frac{\partial h}{n \partial \theta_k}\right|_{\theta} = \left.\frac{\partial h}{n \partial \theta_k}\right|_{\theta_k} + (\bar{\theta} - \theta_k) \left[\frac{\partial^2 h}{n \partial \theta_k^2}\right]_{\theta_k + t(\bar{\theta} - \theta_k)} dt
= \left.\frac{\partial h}{n \partial \theta_k}\right|_{\theta_k} + (\bar{\theta} - \theta_k) \left[\frac{\partial^2 h}{n \partial \theta_k^2}\right]_{\theta_k} + (\bar{\theta} - \theta_k) \int_0^1 \left[\frac{\partial^2 h}{n \partial \theta_k^2}\right]_{\theta_k + t(\bar{\theta} - \theta_k)} - \left.\frac{\partial^2 h}{n \partial \theta_k^2}\right|_{\theta_k} dt
=: \frac{1}{n} \sum_{m=1}^n X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m + (\bar{\theta} - \theta_k) \left.\frac{\partial^2 h}{n \partial \theta_k^2}\right|_{\theta_k} + (\bar{\theta} - \theta_k) H_n.
\]
That is
\[
\frac{\partial h}{n \partial \theta_k} \left[ \frac{\partial^2 h}{n \partial \theta_k^2} \right]^{-1} = (\bar{\theta} - \theta_k) + \left\{ \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m + (\bar{\theta} - \theta_k)H_n \right\} \left[ \frac{\partial^2 h}{n \partial \theta_k^2} \right]^{-1}
\]
\[
= (\bar{\theta} - \theta_k) - T(\bar{\theta} - \theta_k)
\]

From (A.21) and (A.20), it follows that
\[
\|H_n\| \leq 2\|\bar{\theta} - \theta_k\|^{\frac{1}{n}}\sum_{m=1}^{n} \|\eta_m\|^{2+\delta}.
\]

Notice \(E\|\eta_k\|^{2+\delta} < \infty\). From the strong law of large numbers, it follows that
\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \|\eta_m\|^{2+\delta} \leq E\|\eta\|^{2+\delta} =: K_2 < \infty \quad \text{a.s.}
\]

Also, it is easily seen that \(\{X_{n,k}(\xi_{n,k} - p_{n,k})\eta_n\}\) is a sequence of martingale differences with \(E\|X_{n,k}(\xi_{n,k} - p_{n,k})\eta_n\|^{2+\delta} < \infty\). From the law of the iterated logarithm for martingales, it follows that
\[
\sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m = O\left(\sqrt{\log \log n}\right) = o(1) \quad \text{a.s.}
\]

So, for any \(\epsilon > 0\) small enough with \(16K_1K_2\epsilon^\delta < 1\), almost surely for \(n\) large enough we have
\[
\|T(\bar{\theta} - \theta_k)\| \leq \left(\frac{\epsilon}{4K_1} + 4K_2\|\bar{\theta} - \theta_k\|^{1+\delta}\right) \cdot 2K_1 < \epsilon \quad \text{whenever } \|\bar{\theta} - \theta_k\| = \epsilon.
\]

By the Rothe fixed point theory, there exists an \(\tilde{\theta}_{n,k}\) with \(\|\tilde{\theta}_{n,k} - \theta_k\| < \epsilon\) such that
\[
T(\tilde{\theta}_{n,k} - \theta_k) = (\tilde{\theta}_{n,k} - \theta_k).
\]

And then
\[
\frac{\partial h}{n \partial \theta_k} \bigg|_{\tilde{\theta}_{n,k}} = 0.
\]

So, \(\tilde{\theta}_{n,k}\) is a solution of the equation (A.19) and is in the \(\epsilon\)-neighborhood of \(\theta_k\). The first step proof is now finished. And we conclude that, for \(n\) large enough, \(\hat{\theta}_{n,k} = \hat{\theta}_{n,k}^*\) and
\[
0 = \frac{\partial h}{n \partial \theta_k} \bigg|_{\hat{\theta}_{n,k}} = \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m + (\hat{\theta}_{n,k} - \theta_k) \frac{\partial^2 h}{n \partial \theta_k^2} \bigg|_{\theta_k} + (\hat{\theta}_{n,k} - \theta_k)H_n.
\]
Notice that (A.23), (A.22) and 
\[ \|H_n\| = O(\|\hat{\theta}_{n,k} - \theta_k\|^\delta) = o(1) \quad \text{a.s.} \]

From (A.24) it follows that 
\[ \hat{\theta}_{n,k} - \theta_k = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.} \]

It follows that the condition (3.7) in Theorem 3.1 is satisfied, and the second step proof is finished.

Now, by (3.7), for some \( \delta' > 0 \),
\[
\mathbb{E}\left[X_{n,k}p_{n,k}q_{n,k}\eta'_n\eta_n | F_{n-1}\right] = \mathbb{E}_\eta[f_k(\hat{\theta}_{n-1}, \eta)p_k(\theta_k, \eta)]q_k(\theta_k, \eta')\eta + o(n^{-\delta'}) = I_k + o(n^{-\delta'}) \quad \text{a.s.}
\]

With the same argument as showing (A.14), we can show that
\[
\frac{\partial^2 h}{\partial \theta_k^2} = -\frac{1}{n} \sum_{m=1}^{n} \left\{ X_{m,k}p_{m,k}q_{m,k}\eta'_m\eta_m - \mathbb{E}[X_{m,k}p_{m,k}q_{m,k}\eta'_m\eta_m | F_{m-1}] \right\}
\]
(A.25)
\[
\frac{1}{n} \sum_{m=1}^{n} \mathbb{E}[X_{m,k}p_{m,k}q_{m,k}\eta'_m\eta_m | F_{m-1}] = -I_k + o(n^{-\delta'}) \quad \text{a.s.}
\]

Putting (3.7) and (A.25) into (A.24) yields
\[
\left(\hat{\theta}_{n,k} - \theta_k\right)I_k
\]
\[
= \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m + \|\hat{\theta}_{n,k} - \theta_k\| \cdot o(n^{-\delta'}) + O(\|\hat{\theta}_{n,k} - \theta_k\|^{1+\delta})
\]
\[
= \frac{1}{n} \sum_{m=1}^{n} X_{m,k}(\xi_{m,k} - p_{m,k})\eta_m + o(n^{-1/2-\tau}) \quad \text{a.s.}
\]
for some \( \tau > 0 \). (A.15) is now proved. \( \square \)

**Proof of Theorem 5.2.** Notice
\[
X_{m,k}\xi_{m,k} - \mathbb{E}[f_k p_k]
\]
\[
= X_{m,k}\xi_{m,k} - \mathbb{E}[X_{m,k}\xi_{m,k} | F_{m-1}] + \mathbb{E}_\eta[f_k(\hat{\theta}_{m-1}, \eta)p_k(\theta, \eta)] - \mathbb{E}[f_k(\theta, \eta)p_k(\theta, \eta)]
\]
\[
= X_{m,k}\xi_{m,k} - \mathbb{E}[X_{m,k}\xi_{m,k} | F_{m-1}] + \sum_{j=1}^{K} (\hat{\theta}_{m-1,j} - \theta_j)\mathbb{E}\left[\frac{\partial f_k}{\partial \theta_j}p_k\right] + o(\|\hat{\theta}_{m-1} - \theta\|^{1+\delta})
\]

31
By (5.2), it follows that

\[ S_n - n\mu_S = \sum_{m=1}^{n} \sum_{k=1}^{K} (X_{m,k}\xi_{m,k} - \mathbb{E}[X_{m,k}\xi_{m,k}|\mathcal{F}_{m-1}]) + \sum_{m=1}^{n} O(\sqrt{\log \log m} / m) \]

\[ = O(\sqrt{n \log \log n}) \text{ a.s.}, \]

according to the law of the iterated logarithm for martingales. On the other hand, by (A.15), it follows that

\[ S_n - n\mu_S = \sum_{m=1}^{n} \sum_{k=1}^{K} (X_{m,k}\xi_{m,k} - \mathbb{E}[X_{m,k}\xi_{m,k}|\mathcal{F}_{m-1}]) + \sum_{m=1}^{n} O(\sqrt{\log \log m} / m) \]

\[ = \sum_{k=1}^{K} M_{n,k} + \sum_{m=1}^{n} \sum_{j=1}^{K} T_{m,j} \sum_{k=1}^{K} \mathbb{E}\left[ \frac{\partial f_k}{\partial \theta_j} p_k \right] + o(n^{1/2-\tau/2}), \]

where

\[ M_{n,k} = \sum_{m=1}^{n} \Delta M_{m,k}, \quad \Delta M_{m,k} = X_{m,k}\xi_{m,k} - \mathbb{E}[X_{m,k}\xi_{m,k}|\mathcal{F}_{m-1}] \]

and

\[ T_{m,j} = \sum_{i=1}^{m} \Delta T_{i,j}, \quad \Delta T_{m,j} = X_{m,j}(\xi_{m,j} - p_j(\theta_j, \eta_m))\eta_m I_j^{-1}. \]

By computing the conditional co-variance of martingale differences \( \Delta M_{m,k}, \Delta T_{m,j}, k, j = 1, \ldots, K \), and applying the weak convergence theorem for martingales, one can obtain

\[ n^{-1/2}(\mathbb{M}_{[nt],1}, \ldots, \mathbb{M}_{[nt],K}; \mathbb{T}_{[nt],1}, \ldots, \mathbb{T}_{[nt],K}) \quad \Rightarrow \quad (\mathbb{W}_{t,1}, \ldots, \mathbb{W}_{t,K}, \mathbb{B}_{t,1}, \ldots, \mathbb{B}_{t,K}), \]

where \((\mathbb{W}_{t,1}, \ldots, \mathbb{W}_{t,K}, \mathbb{B}_{t,1}, \ldots, \mathbb{B}_{t,K})\) is a multi-dimensional Wiener process with

\[ \mathbb{E}[\mathbb{W}_{t,k}^2] = t \left( \mathbb{E}[f_k p_k] - (\mathbb{E}[f_k p_k])^2 \right), \quad \mathbb{E}[\mathbb{W}_{t,k}\mathbb{W}_{t,j}] = -\tau t \mathbb{E}[f_k p_k] \mathbb{E}[f_j p_j], \ k \neq j; \]

\[ \mathbb{E}[\mathbb{B}_{t,k}^2] = t I_k^{-1}, \quad \mathbb{E}[\mathbb{B}_{t,k}\mathbb{B}_{t,j}] = 0 \ k \neq j; \]

\[ \mathbb{E}[\mathbb{W}_{t,k}\mathbb{B}_{t,k}] = t \mathbb{E}[f_k p_k q_k I_k^{-1}], \quad \mathbb{E}[\mathbb{W}_{t,k}\mathbb{B}_{t,j}] = 0 \ k \neq j. \]

So,

\[ n^{-1/2}(S_{[nt]} - [nt]\mu_S) \Rightarrow \sum_{k=1}^{K} \mathbb{W}_{t,k} + \sum_{j=1}^{K} \int_{0}^{t} \mathbb{B}_{x,j} dx \sum_{k=1}^{K} \mathbb{E}\left[ \frac{\partial f_k}{\partial \theta_j} p_k \right]. \]
It is easily checked that the variance of the above Gaussian process is \( t\sigma^2_S \). The theorem is now proved.

**Proof of Theorems 2.1 and 2.2.** Notice

\[
\frac{\partial f_A}{\partial \theta_A} = -\frac{\partial f_B}{\partial \theta_A} = f_AF_B\eta \quad \text{and} \quad \frac{\partial f_A}{\partial \theta_B} = -\frac{\partial f_B}{\partial \theta_B} = -f_AF_B\eta.
\]

The results follow directly from Theorems 5.1 and 5.2 via some elementary calculations. □

**Proof of Proposition 2.1.** First, we have

\[
\frac{\partial f_A^*}{\partial p_A} = -\frac{\partial f_B^*}{\partial p_A} = \frac{f_A^*f_B^*}{p_A^*q_A^*} \quad \text{and} \quad \frac{\partial f_A^*}{\partial p_B^*} = \frac{f_A^*}{p_B^*q_B^*}.
\]

and so,

\[
\sum_{k=A}^B \frac{\partial f_k^*}{\partial p_j^*} p_k^* = (p_A^* - p_B^*) \frac{\partial f_A^*}{\partial p_j^*}, \quad j = A, B
\]

and

\[
p_A^*q_A^* + p_B^*q_B^* = 0.
\]

Notice that the maximum likelihood estimator of \( p_k^* \) is \( \hat{p}_{m,k} = \sum_{l=1}^m X_{l,k}\xi_{l,k}/N_{m,k}, \quad k = A, B \). Write \( \hat{p}_m^* = (\hat{p}_{m,A}^*, \hat{p}_{m,B}^*) \). Proposition 2.1 (a) is a special case of Theorem 2.1 of Hu and Zhang (2004) (with \( \lambda = 0, \gamma = 1 \)). For proving Proposition 2.1 (c), notice for \( k = A, B \),

\[
X_{m,k}\xi_{m,k} - f_k^* p_k^*
\]

\[
= X_{m,k}\xi_{m,k} - E[X_{m,k}\xi_{m,k}|F_{m-1}] + [f_k^*(\hat{p}_{m-1}^*) - f_k^*(p^*)]p_k^*
\]

\[
= X_{m,k}\xi_{m,k} - E[X_{m,k}\xi_{m,k}|F_{m-1}] + \sum_{j=A}^B (p_k^* - p_j^*) \frac{\partial f_j^*}{\partial p_j^*} p_k^* + o(||\hat{p}_{m-1}^* - p^*||^{1+\delta})
\]

\[
= X_{m,k}\xi_{m,k} - E[X_{m,k}\xi_{m,k}|F_{m-1}] + \sum_{j=A}^B \frac{T_{m,j}^*}{m} \frac{\partial f_j^*}{\partial p_j^*} p_k^* + o(n^{1/2}) \quad \text{a.s.,}
\]

where

\[
T_{m,j}^* = \sum_{l=1}^m \Delta T_{l,j}^* = \sum_{l=1}^m \frac{1}{f_j^*} X_{l,j}(\xi_{l,j} - p_j^*).
\]

Using the similar argument in proving Theorem 5.2, we obtain

\[
S_n - n\mu_S^* = O\left(\sqrt{n \log \log n}\right) \quad \text{and} \quad \sqrt{n}(S_N/n - \mu_S) \overset{D}{\to} N(0, \sigma^2),
\]

33
with
\[ \sigma^2 = \mu_\delta^2 - \mu_\theta^2 + 2 \sum_{j,k=1}^B p_j^* q_k^* \frac{\partial f_k}{\partial p_j^*} p_k^* + 2 \sum_{j=1}^B \frac{p_j^* q_j^*}{f_j^*} \left( \sum_{k=1}^B \frac{\partial f_k^*}{\partial p_j^*} p_k^* \right)^2 = \sigma_\delta^2. \]

At last, we show Proposition 2.1 (b). Notice
\[ \zeta_{m,A}^*(y) = \sum_{m=1}^n \Delta \zeta_{m,A}^*(y) := \sum_{m=1}^n (X_{m,A} - f_A^*) I\{\eta_m = y\} \]
\[ = \sum_{m=1}^n \left( \Delta \zeta_{m,A}^*(y) - \mathbb{E}[\Delta \zeta_{m,A}^*(y)|F_{m-1}] \right) + \sum_{m=1}^n \left( f_A^*(p^*_{m-1}) - f(p^*) \right) P(\eta_m = y) \]
\[ = \sum_{m=1}^n \Delta M_{m,A}^*(y) + \sum_{m=1}^n \sum_{j=1}^B \frac{1}{m} T_m^j \frac{\partial f_A^*}{\partial p_j^*} P(\eta = y) + o(n^{1/2}) \text{ a.s.}, \]
where \( \Delta M_{m,k}^*(y) = \Delta \zeta_{m,k}^*(y) - \mathbb{E}[\Delta \zeta_{m,k}^*(y)|F_{m-1}] \). It is easily checked the multi-dimensional martingale sequence \( \{(M_{n,A}^*(y), T_{n,A}^*, T_{n,B}^*)\} \) satisfying
\[ \mathbb{E}[\Delta M_{n,A}^*(y) \Delta T_{n,j}^*|F_{n-1}] = \frac{f_A^*}{f_j^*} \mathbb{E}[X_{n,j}(\xi_{n,j}^* - p_j^*) I\{\eta_n = y\}|F_{n-1}] \]
\[ = -\frac{f_A^*}{f_j^*} \mathbb{E}[X_{n,j}(\xi_{n,j}^* - p_j^*) I\{\eta_n = y\}] \]
\[ \quad - f_A^*(p_{j|y}^* - p_j^*) P(\eta = y) = \mathbb{E}[W_1^*(y)] B_{1,j}^* \text{ in } L_1 \quad j = A, B, \]
\[ \mathbb{E}[\Delta M_{n,A}^*(y)^2|F_{n-1}] \rightarrow f_A^*(1 - f_A^*) P(\eta = y) = \mathbb{E}[W_1^*(y)^2] \text{ in } L_1, \]
\[ \mathbb{E}[\Delta T_{n,j}^*|F_{n-1}] \rightarrow p_j^* q_j^* \quad \text{ in } L_1 \quad j = A, B \]
and
\[ \mathbb{E}[\Delta T_{n,k}^* \Delta T_{n,j}^*|F_{n-1}] = 0 = \mathbb{E}[B_{1,k}^* B_{1,j}^*] \quad k \neq j \]
where \( (W_t^*(y), B_{t,A}^*, B_{t,B}^*) \) is a 3-dimensional Wiener process. It follows that
\[ n^{-1/2} \zeta_{n,A}^*(y) \Rightarrow W_t^*(y) + \sum_{j=A}^B \int_0^t B_{s,j}^* \frac{\partial f_A^*}{\partial p_j^*} P(\eta = y) \, ds \]
By computing the co-variance matrix of the above Gaussian process, it follows that
\[ n^{-1/2} \zeta_{n,A}^* \overset{D}{\Rightarrow} N(0, \sigma^2), \]
where
\[ \sigma^2 = f_A^*(1 - f_A^*) P(\eta = y) - 2 f_A^* \sum_{j=A}^B (p_{j|y}^* - p_j^*) \frac{\partial f_A^*}{\partial p_j^*} (P(\eta = y))^2 \]
\[ + 2 \sum_{j=A}^B \left( \frac{\partial f_A^*}{\partial p_j^*} \right)^2 \frac{p_j^* q_j^*}{f_j^*} (P(\eta = y))^2 = (\sigma_{A|y})^2 P(\eta = y). \]
Then the theorem is proved by noticing
\[ \sqrt{N_n(y)} \left( \frac{N_{A|y}}{N_n(y)} - f^*_A \right) = \sqrt{n} \frac{\zeta^*_n A(y)}{\sqrt{N_n(y)}} \quad \text{and} \quad \frac{N_n(y)}{n} \rightarrow P(\eta = y) \quad a.s. \] □

Acknowledgments. Professor Hu is also affiliated with the Division of Biostatistics and Epidemiology, Department of Health Evaluation Sciences, School of Medicine, University of Virginia. Part of this research was done while Professor Hu and Professor Zhang visiting Department of Statistics and Actuarial Science, the University of Hong Kong and Department of Statistics, the Chinese University of Hong Kong. They thank both department for their hospitality and support.

REFERENCES


