Minimum Hellinger Distance Estimation for Randomized Play the Winner Design

An-lin Cheng
Department of Biostatistics
Yale University
New Haven, CT 06520-8034

Anand N. Vidyashankar*
Department of Statistical Science
Cornell University
Ithaca, NY 14850-4201

Abstract

Response-adaptive designs in clinical trials incorporate information from prior patient responses in order to assign better performing treatments to the future patients of a clinical study. An example of a response adaptive design that has received much attention in recent years is the Randomized Play the Winner Design (RPWD). Beran (1977) investigated the problem of minimum Hellinger distance procedure (MHDP) for continuous data and showed that minimum Hellinger distance estimator (MHDE) of a finite dimensional parameter is as efficient as the MLE (Maximum Likelihood Estimator) under a true model assumption. This paper develops minimum Hellinger distance methodology for data generated using randomized play the winner design (RPWD). A new algorithm using the Monte Carlo approximation to the estimating eqaution is proposed. Consistency and asymptotic normality of the estimators are established and the robustness and small sample performance of the estimators are illustrated using simulations. The methodology when applied to the clinical trial data conducted by Eli-Lilly and Company, brings out the treatment effect in one of the strata using the frequentist techniques compared to the Bayesian arguemnt of Tamura *et al.* .

---

* Research Supported in part by a grant from NSF DMS 000-03-07057 and also by grants from the NDCHealth Corporation

Key Words: Response Adaptive Randomization, Hellinger distance, Monte-Carlo Approximation, Influence Function, Clinical Trial,

AMS 1991 *subject Classification:* 60J80 60F10

Short title: MHDE for RPWD

1 Introduction

In a typical clinical trial information on several variables are collected. The data includes information on several primary and secondary varaiables, usually called the end-points of the study. For example, consider a clinical trial investigating the efficacy of a drug in lowering the cholesterol levels. In such a trial the primary variables would typically be cholesterol levels, blood pressure, height, diet, life style, and body weight. There are certain secondary variables on which the information is typically collected. These could include information on the genetic components and other heriditary related information. Ofcourse as time evolves, the differences between primary and secondary endpoints may disappear and may even juxtapose.

Response adaptive designs tend to skew patient allocation towards a treatment performing "better" (with respect to one important clinical variable) during the course of a clinical trial. This leads to imbalances in the number of subjects allocated to various treatment groups. Analysis of data resulting from such designed experiments is complicated due to the fact that the sample sizes are dependent random variables. It is customary in a pharmaceutical setting to attempt a parametric modeling and analyze the data using the maximum likelihood methodology (MLM). However, in situations described above, it is not feasible to perform a good parametric analysis of the data, using the maximum likelihood procedure(MLP), owing to the fact that there may not be enough power to verify hypothesis concerning the distributions. Furthermore, since the MLP operates optimistically (by maximizing the likelihood), the resulting inference could be misleading especially if the assumed parametric hypothesis is invalid. To overcome these issues, it may be convenient to rather adopt a pessimistic approach and seek a methodology that would result in "similar" inferences akin to MLP when the assumptions concerning the model hold.

Minimum Hellinger distance Procedure (MHDP), first studied in detail by Beran (1977), attains the dual goal of robustness and efficiency at the true model. In the analysis of a data set from a clinical trial conducted by Eli-Lilly, we discovered a treatment effect in one of the strata using the MHDP and the MLP. Furthermore, the confidence interval constructed using the MHDP was "tighter" and provided consistent inference while those produced by the MLP did not indicate the presence of a "treatment effect" in any starata. Motivated by this inconsistency of the MLP, we sought to understand, from an asymptotic and small sample perspective, if MHDP yields results that are consistent with MLP if the assumed parametric hypotheses are valid. For these reasons, we formally develop the MHDP for analysis of continuous variables obtained from a clinical experiment performed using randomized play the winner design (RPWD). We now move on to describe the RPWD.

1.1 Randomized Play the Winner Design

The RPWD is a method of assigning patients to intervention in clinical trials. This method was inspired from the play the winner rule, originally formulated by Zelen (1969). The play the winner rule can be described as follows: A success on a particular treatment generates a future trial on the same treatment with a new patient. A failure on a treatment generates a future trial on the alternate treatment with a new patient. If the response is unavailable, the subject is allocated using equal probability amongst all treatments. When the patient accrual is rapid and the response is delayed, play the winner rule leads to a complete randomization there by making the adaptation irrelevant. To overcome these difficulties, Wei and Durham (1978) introduced Randomized Play the Winner Design. This design can be described, using an urn model, as follows: Suppose we want to assign patients to two different treatments (treatments 1 and 2). We would start this procedure with an urn containing $n_1(0)$ balls of type 1 and $n_2(0)$ balls of type 2, corresponding to treatments A and B respectively. Once a patient arrives for treatment assignment, a ball is drawn randomly from the urn and returned to the urn. The patient is assigned to a treatment according to the type of ball. When the patient's response is available the urn is updated as follows: if the response is a success on treatment 1 or a failure on treatment 2, then α type 1 balls are added to the urn; however if the response is failure on treatment 1 or a success on treatment 2, then α type 2 balls are added to the urn. The process is repeated until all patients have been assigned to one of the treatments.

Let n denote the total number of patients, N_1 the total number of patients receiving treatment 1 and $n_2(= n - N_1)$ the total number of patients receiving treatment 2. Let $p_1 > 0$ and $p_2 > 0$ denote the probability of success when receiving treatments 1 and 2 respectively. An important question concerns the composition of the urn after n updates to the urn. Athreya and Karlin (1968) show that

$$\frac{N_i}{n} \stackrel{a.s.}{\rightarrow} \pi_i \; as \; n \rightarrow \infty, \tag{1.1}$$

where $\pi_1 = \frac{q_2}{(q_1+q_2)}$ and $q_1 = 1 - p_1, \quad q_2 = 1 - p_2$. Since $\pi_i > 0$, it also follows that

$$\frac{N_i}{n} I_{N_i \geq 1} \stackrel{a.s.}{\rightarrow} \pi_i \; as \; n \rightarrow \infty. \tag{1.2}$$

It turns out that the above results remains valid when one assumes a probability model for the delay mechanism and the arrival process. This was established recently by Bai, Hu and Rosenberger (2002). Some of the statistical questions concerning the RPWD involve estimation and hypothesis testing concerning p_1 and p_2. Rosenberger, Flournoy and Durham (1997) studied maximum likelihood estimation for the parameters of the RPWD. Wei (1988) studied the permutation test for comparing the parameters of the RPWD.

Our theoretical description of the properties of the MHDE involves studying the behavior of $E(\frac{n}{N_i} : N_i \geq 1)$ as $n \to \infty$. Our original proof of the proposition contained a gap that involved a dominated convergence argument. It turns of that, such a method was more subtle than we anticipated and provide the following proof by synthesizing an idea of the anonymous referee. The situation $\alpha = 1$ and $\alpha \geq 2$ invove different arguments with The method of proof involves a large deviation estimate and the imbedding of the urn scheme into a two-type Galton Watson process. We now state our proposition:

Proposition 1.1 $\lim_{n\to\infty} E(\frac{n}{N_i} I_{N_i \geq 1})^k = \pi_i^{-k}$

Proof. Fix an $\epsilon > 0$ and let $A_1 = \{N_i \leq n(\pi_i - \epsilon)\}$, $A_2 = \{n(\pi - \epsilon) \leq N_i \leq n(\pi_i + \epsilon)\}$ and $A_3 = \{N_i > n(\pi_i + \epsilon)\}$. Then,

$$E(\frac{n}{N_i} I_{N_i \geq 1})^k = \sum_{r=1}^{3} E((\frac{n}{N_i} I_{N_i \geq 1})^k : A_r).$$

It is easy to see using (1.2) that the above expectation on A_3 converges to 0 as $n \to \infty$. Also using the upper and lower bounds on A_2, one can show that the expectation on A_2 converges to the desired limit stated in the proposition. Thus to complete the proof, we only have to establish that the expectation on A_1 converges to 0 as $n \to \infty$. To this end, note that

$$\begin{aligned} E((\frac{n}{N_i} I_{N_i \geq 1})^k : A_2) &\leq \sum_{r=1}^{n-1} E((\frac{n}{N_i} I_{N_i \geq 1})^k : r(\pi_i - \epsilon) \leq N_i \leq (r+1)(\pi_i + \epsilon)) \\ &\leq n^k (\sum_{r=1}^{n-1} r^{-k}) P(\frac{N_i}{n} \leq \pi_i - \epsilon) \end{aligned} \tag{1.3}$$

Note that, using the embedding technique and the generating function given on page 1805 equation (15) of Athreya and Karlin(1968) and large deviation techniques (which uses functional iteration) developed in Vidyashankar(1994), and Athreya and Vidyashankar (1995), one can show that there exist universal constants $C_1(\pi_i, \epsilon) > 0$ and $C_2(\pi_i, \epsilon) > 0$ such that

$$\begin{aligned} P(\frac{N_i}{n} \leq \pi_i - \epsilon) &\leq P(\frac{Z_{i\tau_n}}{Z_{1\tau_n} + Z_{2\tau_n}} \leq \pi_i - \epsilon) \\ &\leq C_1 exp(-C_2 2^{\log n}) \end{aligned} \tag{1.4}$$

where $(Z_{1\tau_n}, Z_{2\tau_n})$ is the population sizes of number of type 1 and type 2 during the n^{th} split of a two-type branching process. The result follows using (1.4) in (1.3). ∎

We now describe the minimum Hellinger distance estimation procedure for the i.i.d. data.

1.2 Minimum Hellinger Distance Estimators

In this section, we will briefly discuss minimum Hellinger distance estimation for continuous i.i.d. data. Let $f(x)$ and $g(x)$ be any two densities; the Hellinger distance between $f(x)$ and

$g(x)$ is defined as the L_2-norm of the difference between square root of density functions , i.e.

$$HD^2(f,g) = ||f(x)^{1/2} - h(x)^{1/2}||_2^2 = \int [(f(x))^{1/2} - (h(x))^{1/2}]^2 dx. \tag{1.5}$$

Let $X_1, X_2, \cdots, X_n$ be i.i.d. real valued random variables with density belonging to a specified parametric family $\{f_\theta : \theta \in \Theta\}$. To motivate the MHDP, replace f by f_θ and h by h_n, a non-parametric estimator of the density. Therefore, the Hellinger distance in our question becomes the distance between the true density (f_θ) and the nonparametric density estimator of the X_i's, which can be expressed as follows:

$$HD_n^2(f_\theta, h_n) = ||f_\theta(x)^{1/2} - h_n(x)^{1/2}||_2^2. \tag{1.6}$$

The Minimum Hellinger distance estimator of θ is defined to be the value $\hat{\theta_n}$ (in the parameter space Θ), if it exists, that minimizes (1.5). Using simple algebra, one can show that

$$HD_n^2(f_\theta, h_n) = 2 - 2\gamma_n(\theta)$$

where

$$\gamma_n(\theta) = \int (f_\theta(x))^{1/2} (h_n(x))^{1/2} dx.$$

Hence finding the minimum Hellinger distance estimator is therefore equivalent to finding the $\hat{\theta_n}$ that maximizes $\gamma_n(\theta)$.

If one chooses

$$h_n(x) = \frac{1}{nc_n} \sum_{j=1}^{n} K\left\{\frac{x - X_j}{c_n}\right\}$$

where $K(\cdot)$ is a kernel density, then it is well-known that (see Devroye (1987)) as $c_n \to 0$, $h_n(x) \xrightarrow{L_1} f_\theta(x)$. This implies that $HD_n^2(f_\theta, h_n) \to 0$. This argument suggests investigating estimators that minimize the Hellinger distance between the nonparametric density estimator and the proposed parametric density.

Beran (1977) has shown that the MHDE is more "robust" than maximum likelihood estimator when data contaminations are present. Furthermore MHDE is known to be asymptotically efficient under a specified parametric family of densities and is minimax robust in a small Hellinger metric neighborhood of the given family (Beran 1977).

RPWD methodology naturally leads to situations where fewer subjects are sometimes allocated to one of the treatment arms. In these situations, it is usually difficult to identify the true distribution of the data under consideration (due to the lack of power) in order to perform adequate parametric inference. Drawing on the results from the i.i.d. literature, it is conceivable that methodology based on MHDE would be more robust than the MLE and perhaps just as efficient as the MLE. For these reasons, in this paper we will under take a systematic development of minimum Hellinger distance procedure for the analysis of data from RPWD.

The rest of the paper is structured as follows: Section 2 develops MHDP for data from RPWD and studies the existence and uniqueness of the MHDE. Section 3 is devoted to a study of the consistency of the minimum hellinger distance estimator(s). Section 4 discusses limit distributions and while section 5 deals with robustness. Section 6 provides a new monte-carlo based computational algorithm for obtaining the MHDE while section 7 describes various simulation results. Section 8 contains some concluding remarks.

We end this section with a description of an analysis of Eli-Lilly data which motivated the entire study.

1.3 A Motivating Data Analysis Example

We now describe the clinical trial conducted by Eli-Lilly and company which partly motivated this paper. This is a multi-center clinical trial comparing fluoxetine to placebo in patients with depressive disorder. It is believed (Kupfer (1976)) that shortened rapid eye movement latency is a marker for endogenous depression. In this trial, patients were stratified into two groups: Patients with normal rapid eye movement latency(REML) and patients with shortened REML. The first six patients within each stratum were assigned by a randomized block design to either fluoxetine or placebo. The trial used two independent urns (for two different strata) to assign the patients. Both urns started with one ball for each type, representing the two treatments. Independent randomized play the winner rules were initiated with the seventh patient within each stratum. There are two primary outcomes: (1) the percentage of patients who exhibited a 50 percent or greater reduction in Hamilton Depression Scale ($HAMD_{17}$) between baseline and final active visit after a minimum of three weeks of therapy, and (2) the reduction in $HAMD_{17}$ between baseline and the final visit. Patients receiving therapy for at least 3 weeks who exhibited a 50 percent or greater reduction in $HAMD_{17}$ were defined to be responders (success in treatment). The time from baseline to final measurement was approximately 8 weeks. The time delay, along with a rapid patient arrival, did not allow an adaptive trial based on the response from final visit. Thus adaptive allocation was based on a surrogate marker to update the urn. The surrogate responder was defined as a patient exhibiting a reduction greater than 50 percent in ($HAMD_{17}$) in two consecutive visits after at least three weeks of therapy. The trial was stopped after 61 patients had responded according to the surrogate criterion. No further surrogate response was obtained for the remaining patients. There were total 89 patients in this trial.

The data related to this trial is available in (23) where shortened REML patients belong to strata 1, normal REML patients belongs to strata 0; treatment is denoted by 1 if the patients were treated with Fluoxetine and 0 if they were treated with placebo. A total of 83 patients have a final response recorded.

The following Table 1.1 presents MHDE and MLE of mean and standard deviations and

their asymptotic confidence intervals for the primary outcome, the difference of $HAMD_{17}$ between the baseline and the final visit for both the strata.

Table 1.1: Analysis results for the Fluoxetine trial data

Strata	Strata=1 Treatment=1	Strata=1 Treatment=0	Strata=0 Treatment=1	Strata=0 Treatment=0
MHDE	-10.88 (5.04)	-3.92 (5.73)	-10.14 (6.51)	-9.59 (5.96)
Asy. CI by MHDE	(-13.09, -8.67)	(-6.37, -1.47)	(-12.92, -7.36)	(-12.14, -7.04)
Length	(4.42)	(4.9)	(5.56)	(5.1)
MLE	-11.20 (5.97)	-5.71 (7.68)	-10.81 (7.13)	-8.62 (6.88)
Asy. CI by MLE	(-13.82, -8.58)	(-8.99, -2.48)	(-13.86, -7.76)	(-11.56, -5.68)
Length	(5.24)	(6.51)	(6.1)	(5.88)

Our next table, Table 1.2 provides the Z- statistics and t-Statistics and the p-values using the estimates from the MHDP and the MLP.

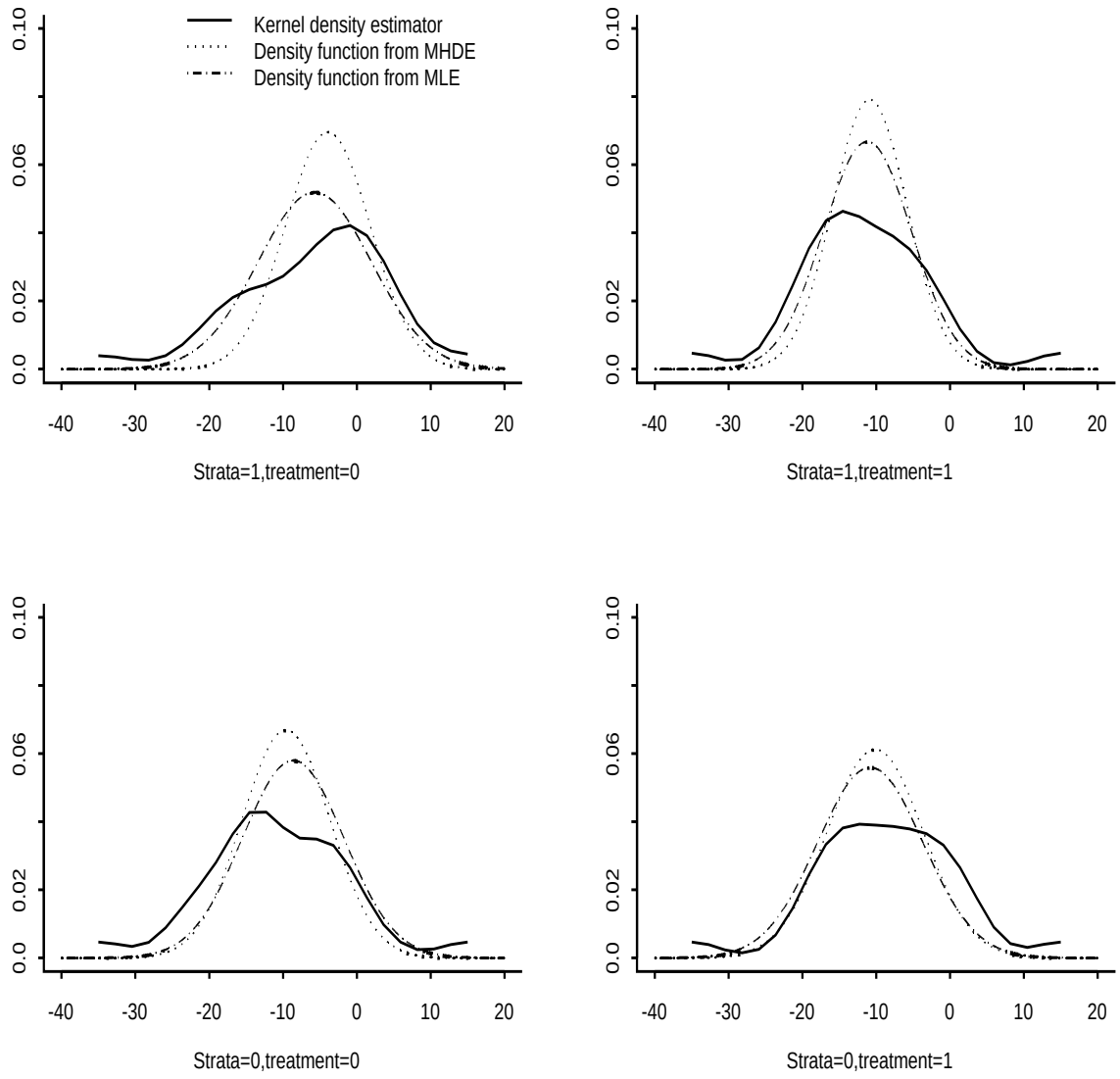
Table 1.2: P value results for the Fluoxetine trial data

Strata	Strata=1 Statistic(p-value)	Strata=0 Statistic(p-value)
MHDE (Z test)	4.1347 (0.00004)	0.3011(0.76332)
MHDE (t test)	13.1914(0)	0.9758(0.1675)
MLE (Z test)	2.5612(0.0104)	0.9885(0.327)
MLE (t test)	2.5441(0.0075)	1.0129(0.1586)

As can be seen the from the above tables, due to the robustness property, the MHDE of the standard deviation are much "smaller" than that of the MLE. Furthermore, the absolute change in $HAMD_{17}$ between the baseline and final visits seem to be substantially higher for treatment 1 than treatment 0 and the change is more pronounced in strata 1 than in strata 0.

We observe that in strata 1, the confidence intervals obtained by the MHDP do not overlap, showing that there could be a treatment effect in strata 1. This feature was not reflected by the MLE. This is further confirmed using other procedures like deviance testing and small sample methods which is subject matter of a different paper. We do reiterate that the results from asymptotic theory could be suspect since the number of observations is less than 20 in each treatment group.

An anonymous referee raised the question if the differences seen in the placebo group suggest that the responses should be modeled as a mixture of normal distributions. The following graphs provide the best fitting normal distributions for each of the strata and the smoothed kernel density estimates. The graphs do not suggest a mixture of normal distributions for the data but we once again reiterate that these are based on very small sample sizes.



2 MHDE for RPWD

In this section we systematically develop the notations, assumptions, and the terminology that will be in force throughout this paper. Let X_{ij} denote the measurement on the j^{th} patient receiving i^{th} treatment. The data from a RPWD experiment can be expressed as:

$$\begin{array}{cccc} X_{1\nu(11)}, & X_{1\nu(12)}, & \cdots, & X_{1\nu(1n_1)} \\ \vdots & & & \\ X_{k\nu(k1)}, & X_{k\nu(k2)}, & \cdots, & X_{k\nu(kn_k)} \end{array}$$

where $\nu(ij) =$ is index of the j^{th} patient receiving treatment i amongst patients $\{1, 2, \cdots, n\}$. Let $\{T_j : j \geq 1\}$ denote the treatment indicators, i.e.

$$T_j = i \text{ if the } j^{\text{th}} \text{ subject receives the } i^{\text{th}} \text{ treatment.}$$

Let us assume henceforth that we have only two treatments. We will make the following assumption on our data:

(A0) For each $i = 1, 2$, conditioned on $\{\nu(ij), j = 1, \ldots n_i\}$, the sequences of responses $\{X_{i\nu(ik)}, k \geq 1\}$ are i.i.d. with distributions F_{X_i}.

The above assumption implies that the methodology developed in this paper can be applied to perform inference on variables that are not correlated with the randomization variable. In a typical clinical trial, there are several such variables and frequently, the randomizing variable is a surrogate response which turns out to be uncorrelated with other primary end points. In all such situations, the methodology of this paper can be employed.

We begin by recalling a theorem of Melfi and Page(see (17)).

Theorem 2.1. Let $\{(X_{1n}, X_{2n} : n \geq 1\}$ be a collection of i.i.d. random vectors with marginal distributions F_{X_1} and F_{X_2} respectively. Let $\mathcal{F}_n$denote a filtration such that (X_{1n}, X_{2n}) is $\mathcal{F}_n$ measurable but independent of $\mathcal{F}_{n-1}$. For each $i = 1, 2$, let $\{\nu(ij), j \geq 1\}$ be a collection of positive, increasing, almost-surely finite random variables such that $\{\nu(ik) = j\} \in \mathcal{F}_{j-1}$. Then

1. $\{X_{1\nu(1k)}, k \geq 1\}$ are i.i.d. with distribution F_{X_1}.
2. $\{X_{2\nu(2k)}, k \geq 1\}$ are i.i.d. with distribution F_{X_2}.
3. the two sequences $\{X_{1\nu(1k)}, k \geq 1\}$ and $\{X_{2\nu(2k)}, k \geq 1\}$ are independent.

Remark 2.2 The above theorem asserts that the allocated sequences inherit the independence and the distributional structure of the original sequence.

Remark 2.3 An anonymous referee has pointed out to us that the proof of (iii) as available in (17) is incorrect. Apparently a correction to (iii) is being made. We note that, we use only (i) and (ii) of Theorem 2.1 in sections 3 and 4 and use the theorem in this section for motivational purposes only.

We now use the above theorem to develop a Hellinger distance based criterion fucntion for RPWD data. We assume that

(A1) $X_{1\nu(1i)} \sim f(\cdot|\theta)$, $X_{2\nu(2j)} \sim g(\cdot|\eta)$, where $\theta \in \Re^p$, $\eta \in \Re^p$ and θ and η and are not functionally dependent, i.e. there does not exist any $h : \Re^p \to \Re^p$ such that $h(\theta) = \eta$.

Remark 2.4 The clinical significance of the above assumption is that there is no information about treatment 1 from subjects receiving treatment 2 and vice-versa.

Remark 2.5 The condition concerning the functional independence can be removed by modelling the dependence between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. This introduces complex notations and technical issues and will be pursued in a different publication.

Before introducing the criterion function, we need few more notations. Let

$$F(\cdot|\ \Xi) = \begin{pmatrix} f(\cdot|\theta) \\ g(\cdot|\eta) \end{pmatrix} \text{ and } H = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \tag{2.7}$$

where $\Xi = (\theta,\ \eta)'$. Let $\Theta_1 \in \Re^p$ denote the parameter space corresponding to θ, $\Theta_2 \in \Re^p$ denote the parameter space corresponding to η and $\Theta = \Theta_1 \times \Theta_2$ denote the parameter space corresponding Ξ. Let

$$vhd(F,\ H) = \begin{pmatrix} HD^2(f(\cdot|\theta), h_1) \\ HD^2(g(\cdot|\eta), h_2) \end{pmatrix} \tag{2.8}$$

denote the vector of squares of Hellinger distances between the components of F and H. Since we will deal with square root of densities, for the sake of compactness of various expressions, we introduce the following notations:

$$s_1(x|\theta) = f^{\frac{1}{2}}(x|\theta) \qquad\qquad s_2(x|\eta) = g^{\frac{1}{2}}(x|\eta). \tag{2.9}$$

Let $\mathcal{G}$ denote the class of densities metrized by the L_1 distance. We define the MHDF to be the functional (possibly multivalued) $T : \mathcal{G} \times \mathcal{G} \to \Theta$ such that

$$\begin{aligned} T(H) &= \arg\min_{\Xi\in\Theta}\{vhd(F(\cdot|\ \Xi),\ H\} && (2.10)\\ &= \arg\max_{\Xi\in\Theta}\Gamma(\Xi,\ H)\} && (2.11) \end{aligned}$$

where

$$\begin{aligned} \Gamma(\Xi,\ H) &= (\gamma_1(\theta),\ \gamma_2(\eta))' \\ &= \left(\int_{\Re} s_1(x|\theta)h_1^{1/2}(x)dx,\ \int_{\Re} s_2(x|\eta)h_2^{1/2}(x)dx\right)'. \end{aligned}$$

In the above equation it should be noted that the argmin of a vector functional is defined component-wise as follows:

$$\arg\min_{\Xi\in\Theta}(.) = \begin{pmatrix} \arg\min_{\theta\in\Theta_1}(.) \\ \arg\min_{\eta\in\Theta_2}(.) \end{pmatrix}.$$

Now, if H_n are the estimators of H based on the data (to be described below) then the MHDE of Ξ is given by

$$T(H_n) = \arg\min_{\xi\in\Theta}\{vhd(F(\cdot|\ \Xi),\ H_n\}. \tag{2.12}$$

We choose for $H_n \equiv (h_{1n}, h_{2n})$, the following Kernel density estimate, *viz.*:

$$\begin{aligned} h_{in}(x) &= \frac{1}{N_i}\sum_{j=1}^{N_i} K\left(\frac{x - X_{i\nu(ij)}}{c_n}\right) \quad i = 1, 2 \\ &= \frac{1}{N_i}\sum_{j\in\mathcal{A}_n(i)} K\left(\frac{x - X_{ij}}{c_n}\right) \quad i = 1, 2 \end{aligned} \tag{2.13}$$

where

$$\mathcal{A}_n(i) = \{1 \leq j \leq n | T_j = i\}. \tag{2.14}$$

Now using (2.14) in the definition of vhd yields the sample version vhd_n given by

$$\begin{aligned} vhd_n(F,\ H) &= (HD_n^2(f(\cdot|\theta), h_{1n}),\ HD_n^2(g(\cdot|\eta), h_{2n}))' \\ &= 2\cdot\mathbf{1} - \Gamma_n(\Xi,\ H) \end{aligned} \tag{2.15}$$

where $\mathbf{1} = (1, 1)'$.

2.1 Existence and Uniqueness

In this section, we will establish the existence and uniqueness of the MHDE, defined through a minimization of (2.12).We will make the following assumptions through out this paper.

(A2) The parameter spaces Θ_1 and Θ_2 are locally compact.

(A3) $f(\cdot|\theta)$ and $g(\cdot|\eta)$ are upper semi-continuous.

Our first theorem shows that under a further weak regularity condition (2.12) exists.

Theorem 2.1. Assume **(A0)-(A3)**. Let $\Theta_K = K_1 \times K_2$, where $K_i \subset \Theta_i$ is compact for all $i = 1, 2$. Let $\Theta^c = \Theta_1^c \times \Theta_2^c$, $\Theta_1^c = K_1^c \cap \Theta_1$, and $\Theta_2^c = K_2^c \cap \Theta_2$. Assume further that

$$\sup_{\Xi\in\Theta^c} \Gamma(\Xi,\ H) < \sup_{\Xi\in\Theta_K} \Gamma(\Xi,\ H) \tag{2.16}$$

and that

$$\Xi_1 \neq \Xi_2 \Rightarrow H(\cdot|\Xi_1) \neq H(\cdot|\Xi_2) \tag{2.17}$$

on a set of positive Lebesgue measure. Then (2.12) exists and is unique.

Proof. The proof of existence involves two steps.

(1) We will show that $vhd(F,\ H)$ is lower semi-continuous.

(2) We will then use (2.16) along with condition (2.17) of the theorem to establish the existence of the minimizers of (2.12).

We begin with (1). Note that

$$vhd(F(\cdot|\Xi),\ H) = 2 \cdot \mathbf{1} - 2\Gamma(\Xi,\ H). \tag{2.18}$$

Under **(A3)**, $\Gamma(\Xi,\ H)$ is upper semi-continuous function. Hence $vhd(F,\ H)$ is lower semi-continuous. Now, since K_1 and K_2 are compact subsets of Θ_1 and Θ_2 respectively, $\Theta_K = K_1 \times K_2$ is also a compact subset of Θ. Hence from the lower semi-continuity of $vhd(.)$ there exists a $\Xi^* \in \mathbf{\Theta}_K$ such that

$$(m_1, m_2)' = vhd(F(\cdot|\Xi^*),\ H) = \inf_{\Xi \in \Theta_K} vhd(F(\cdot|\Xi),\ H). \tag{2.19}$$

Hence using (2.18)

$$\Gamma(\Xi^*,\ H) = \frac{1}{2}(2 - m_1,\ 2 - m_2)'.$$

Now, using condition (2.16), we have that for all $\Xi \notin \Theta_K$, $\Gamma(\Xi,\ H) < \frac{1}{2}(2 - m_1,\ 2 - m_2)'$. Hence Ξ^* minimizes $vhd(\cdot|\Xi)$ on Θ.

We next prove the uniqueness of $T(F(\cdot|\Xi))$. Asumption (2.17) of the theorem implies identifiabilty. Hence, by applying Beran's (1977) Theorem 1 to the components of

$$vhd(F(\Xi),\ F(\Xi_0)) = \begin{pmatrix} HD^2(f(\cdot|\theta), f(\cdot|\theta_0)) \\ HD^2(g(\cdot|\eta), g(\cdot|\eta_0)) \end{pmatrix}$$

it follows that $vhd(.,.)$ is minimized at $\theta = \theta_0$ and $\eta = \eta_0$ uniquely . ∎

3 Consistency of MHDE for RPWD

In this section we establish the consistency of the MHDE defined in (2.9). The main technical tool involves (i) establishing the continuity of the VHDF and (ii) establishing the L_1 consistency of the $H_n(.)$ defined in (2.14). If one assumes that the parameter space is compact, then (i) and (ii) imply consistency follows from Beran's arguments. We will show that under the regularity conditions assumed in Theorem 3.3 below, one can get consistency without assuming compactness. We begin by studing the L_1 convergence of Kerenl density estimators of RPWD.

3.1 L_1-convergence of Kernel Density Estimators of RPWD

We begin by considering the L_1 consistency of $H_n(.)$. We recall that the kernel density estimators of h_1 and h_2 (the densities of responses for treatment 1 and treatment 2 respectively) are given by

$$h_{in}(x) = (N_i c_n)^{-1} \sum_{j \in \mathcal{A}_n(i)} K\left(\frac{x - X_{ij}}{c_n}\right), \quad i = 1, 2.$$

where $\mathcal{A}_n(i)$ is defined in (2.14). Our first theorem establishes the strong pointwise consistency and strong L_1 consistency of $h_{i,n}(\cdot)$ and $E(h_{i,n}(\cdot))$.

Proposition 3.1. Assume that $c_n \to 0$ and $nc_n \to \infty$ as $n \to \infty$. Then for almost all x (with respect to the Lebesgue Measure)

$$\lim_{n\to\infty} h_{in}(x) = h_i(x) \quad \text{a.s.}, \tag{3.20}$$

and

$$\lim_{n\to\infty} E(h_{in}(x)) = h_i(x). \tag{3.21}$$

Furthermore,

$$\lim_{n\to\infty} \int_{\Re} |h_{in}(x) - h_i(x)| = 0 \quad \text{a.s.}, \tag{3.22}$$

$$\tag{3.23}$$

and

$$\lim_{n\to\infty} \int_{\Re} |E(h_{in}(x)) - h_i(x)| = 0. \tag{3.24}$$

Proof. By Melfi's Theorem, $\{X_{i,\nu(ij)}, j \in \mathcal{A}(i)\}$ are i.i.d. random variables. Since $\frac{N_i}{n}$ converges to $\pi_i > 0$, it follows that $N_i(n)c_n \to \infty$ as $n \to \infty$. Hence, by Theorem 1 of Devroye (1987), (3.20) follows. (3.21) is a consequence of Glick's Theorem. We next calculate

$$\begin{aligned} E\left(K\left(\frac{x - X_{i\nu(ij)}}{c_n}\right)\right) &= \int K\left(\frac{x-y}{c_n}\right) h_i(y)dy \\ &= c_n \int K(t) h_i(x + tc_n)dt \end{aligned} \tag{3.25}$$

Now, conditioning on the treatment assignment, and using (3.25) and assumption **(A0)**

$$E(h_{in}(x)) = \int K(t) h_i(x + tc_n)dt. \tag{3.26}$$

Thus, to complete the proof of (3.22) we need to show that (3.26) converges to $h_i(x)$. Now

$$|E(h_{in}(x)) - h_i(x)| \;\leq\; \int K(t)|h_i(x + tc_n) - h_i(x)|dt. \tag{3.27}$$

By the bounded convergence theorem, right hand side of (3.27) converges to 0 as $n \to \infty$ yielding (2.17). Finally, by integrating (3.27) and interchanging the order of integration (using Tonelli's theorem), it follows again by the bounded convergence theorem that

$$\lim_{n\to\infty} \int |E(h_{in}(x)) - h_i(x)| = 0$$

yielding (3.24). ∎

Remark 3.2. Since convergence in L_1 implies convergence in the Hellinger metric, it follows that

$$\lim_{n\to\infty} HD(h_{in},\ h_i) = 0 \quad \text{a.s. for all} \quad i = 1, 2. \tag{3.28}$$

3.2 Continuity and Consistency of the MHDF

In this section, we study the consistency of the MHDE *via* the continuity of the MHDF defined in (2.12) Recall that $\mathcal{G}$ is the class of densities metrized by the L_1 distance and $T : \mathcal{G} \times \mathcal{G} \to \boldsymbol{\Theta}$ is defined as

$$T(H) = \arg\max_{\Xi\in\Theta} \Gamma(\Xi|H).$$

Our first result establishes the continuity of T. Assume that **(A1)-(A3)** and conditions of Theorem 2.1.1. hold.

Proposition 3.2. Assume further **(A0)-(A3)** that $T(\boldsymbol{H})$ is unique. Then T is continuous, *i.e.* if $h_{1n} \xrightarrow{L_1} h$, $h_{2n} \xrightarrow{L_1} h_2$, then

$$\lim_{n\to\infty} T(H_n) = T(H). \tag{3.29}$$

Proof. Let $h_{1n} \xrightarrow{L_1} h_1$ and $h_{2n} \xrightarrow{L_1} h_2$. By Theorem 2.1, there exists $\Xi_n \in \Theta$ such that $T(H_n) = \Xi_n$. Furthermore, since the minimizers exist inside a compact set, $\{\Xi_n : n \geq 1\}$ is bounded sequence. By another application of Theorem 2.1, there exist $\Xi \in \Theta_K \subset \Theta$ such that

$$T(H) = \Xi.$$

Thus, to prove (3.29) it is enough to show that

$$\Xi_n \to \Xi_0. \tag{3.30}$$

We now show that is sufficient to prove that

$$\lim_{n\to\infty} d_n = \mathbf{0} \tag{3.31}$$

where

$$d_n \equiv \sup_{\Xi \in \Theta} |vhd(F(\cdot|\Xi),\ H_n) - vhd(F(\cdot|\Xi),\ H)|. \tag{3.32}$$

To this end, suppose (3.31) holds and (3.30) does not hold. By the boundedness of $\{\Xi_n : n \geq 1\}$ and the compactness of Θ_K, there exists $\Xi_* (\neq \Xi_0) \in \Theta_K$ and a subsequence n_k such that

$$\Xi_{n_k} \to \Xi_*. \tag{3.33}$$

Hence by (3.31)

$$vhd(F(\cdot|\Xi_{n_k}),\ H_{n_k}) \to vhd(F(\cdot|\Xi_*),\ H). \tag{3.34}$$

This implies that

$$vhd(F(\cdot|\Xi_*),\ H) = vhd(F(\cdot|\Xi),\ H)$$

contradicting the uniqueness of $T(H)$. Now we show that (3.31) holds. By the Cauchy-Schwarz inequality, the components of d_n are bounded above by $HD(h_{1n}, h_1)$ and $HD(h_{2n}, h_2)$ respectively. Now (3.31) follows from Remark 3.2 using (3.28) ∎

Now, using the compactness of K_1 and K_2 and using Theorem 2.1, Proposition 3.1 and Proposition 3.2 we get the strong consistency of the MHDE for RPWD. We state this as a Theorem.

Theorem 3.3. Assume **(A1)-(A3)** holds and that $T(H)$ is unique. Then, the sequence of MHDE defined in (2.12) converges a.s. to $T(H)$. ∎

In the next section, we establish the joint asymptotic normality of the MHDE of Ξ_n. The properties of the kernel density estimate$K(.)$ will play an important role in the proof of asymptotic normality of the MHDE. We will state these conditions and they will be in force throughout the next section.

(K1) $K(.)$ is symmetric about 0 with compact support. We will denote the support of $K(.)$ by $Supp(K)$.

(K2) The window-width c_n satisfies the following: $c_n \to 0, \quad nc_n^2 \to 0, \quad$ and $n^{\frac{1}{2}} c_n \to \infty$.

4 Joint Asymptotic Normality of MHDE of Ξ_0

In this section, we deal with the joint asymptotic normality Ξ_n. We will assume throughout this section that the conditions **(A1)-(A2)** holds. We need the following further differentiability conditions on the families of densities, $\{f(\cdot|\theta) : \theta \in \Theta_1\}$ and $\{g(\cdot|\eta) : \eta \in \Theta_2\}$.

(D1) $f(\cdot|\theta)$ and $g(\cdot|\eta)$ are twice continuously differentiable functions of θ and η.

(D2) Assume further that $||\nabla s_1(\cdot|\theta)||_2$ and $||\nabla s_2(\cdot|\eta)||_2$ are continuous and bounded.

To state our conditions for asymptotic normality precisely, we introduce the vector score functions, *viz.*

$$u_1(\cdot|\theta) \equiv \nabla f(\cdot|\theta) f^{-1}(\cdot|\theta) \quad \text{and} \quad u_2(\cdot|\theta) \equiv \nabla g(\cdot|\eta) f^{-1}(\cdot|\eta). \tag{4.35}$$

Hence,

$$\dot{s}_1\ (\cdot|\theta) = \frac{1}{2} u_1(\cdot|\theta)\ s_1(\cdot|\theta), \quad \text{and} \quad \dot{s}_2\ (\cdot|\eta) = \frac{1}{2} u_2(\cdot|\eta)\ s_1(\cdot|\eta). \tag{4.36}$$

Furthermore, the kl^{th} element of the matrix of second partials of $s_1(\cdot|\theta)$ and $s_2(\cdot|\eta)$ are given by

$$\ddot{s}_{1kl}\ (\cdot|\theta) = \frac{1}{2}\ \dot{u}_{1kl}\ (\cdot|\theta) s_1(\cdot|\theta) + \frac{1}{4}(u_1(\cdot|\theta) u_1'(\cdot|\theta))_{kl} s_1(\cdot|\theta) \tag{4.37}$$

and

$$\ddot{s}_{2kl}\ (\cdot|\eta) = \frac{1}{2}\ \dot{u}_{2kl}\ (\cdot|\eta) s_2(\cdot|\theta) + \frac{1}{4}(u_2(\cdot|\eta) u_2'(\cdot|\eta))_{kl} s_2(\cdot|\eta) \tag{4.38}$$

respectively. Note that $\dot{u}_{1kl}\ (\cdot|\theta)$ represents the kl^{th} element of the matrix $\dot{u}_1\ (\cdot|\theta)$ and $\dot{u}_{2kl}\ (\cdot|\eta)$ the kl^{th} element of the matrix $\dot{u}_2\ (\cdot|\eta)$. We will have occasion to use the Fisher information matrices $I_1(\theta_0)$ and $I_2\eta_0$, defined to be

$$I_1(\theta_0) = \int_{\mathcal{R}} u_1(x|\theta_0) u_1'(x|\theta_0) f(x|\theta_0) dx \tag{4.39}$$

and

$$I_2(\eta_0) = \int_{\mathcal{R}} u_2(x|\eta_0) u_2'(x|\eta_0) g(x|\eta_0) dx. \tag{4.40}$$

Using the (**D1**) and (**D2**) and partially differentiating (2.15) with respect to $\boldsymbol{\Xi}$ we get

$$\nabla v h d_n(\Xi) = 0 \tag{4.41}$$

Let Ξ_n be the solution to (4.41). Now applying one term Taylor expansion of (4.41) we get

$$\nabla v h d_n(\Xi_0) = \nabla v h d_n(\Xi_n) + D_n(\Xi_n^*)(\Xi_n - \Xi_0) \tag{4.42}$$

where (using **(A1)**) $D_n(.)$ is given by

$$D_n(\Xi_n^*) = Diag(D_{1n}(\theta_n^*), \quad D_{2n}(\eta_n^*)). \tag{4.43}$$

and $\Xi_n^* = (\theta_n^*, \eta_n^*)' \in U_n(\theta_0) \times V_n(\eta_0)$ where,

$$U_n(\theta_0) = \{\theta | \theta = t\theta_0 + (1-t)\theta_n\}, \tag{4.44}$$

$$V_n(\eta_0) = \{\eta | \eta = t\eta_0 + (1-t)\eta_n\}, \tag{4.45}$$

$$D_{1n}(\theta) = \frac{1}{2} \int_{\mathcal{R}} \dot{u}_1\ (\cdot|\theta) s_1(\theta) h_{1n}^{\frac{1}{2}}(x) dx + \frac{1}{4} \int_{\mathcal{R}} u_1(\cdot|\theta) u_1'(\cdot|\theta)\ \dot{s}_1\ (\theta) h_{1n}^{\frac{1}{2}}(x) dx, \tag{4.46}$$

and

$$D_{2n}(\theta) = \frac{1}{2}\int_{\mathcal{R}} \dot{u}_2\ (\cdot|\eta)s_2(\eta)h_{2n}^{\frac{1}{2}}(x)dx + \frac{1}{4}\int_{\mathcal{R}} u_2(\cdot|\eta)u_2'(\cdot|\eta)\ \dot{s}_1\ (\eta)h_{1n}^{\frac{1}{2}}(x)dx. \tag{4.47}$$

Thus,

$$(\Xi_n - \Xi_0)' = D_n^{-1}(\Xi_n^*)\nabla vhd_n(\Xi_0). \tag{4.48}$$

Now using the definition of $vhd(.)$, the above simplifies to

$$(\Xi_n - \Xi_0)' = (D_{1n}^{-1}(\theta_n^*)\ \nabla vhd_{1n}(\theta_0), \quad D_{2n}^{-1}(\eta_n^*)\ \nabla vhd_{2n}(\eta_0))'. \tag{4.49}$$

Writing down the expressions and simplifying (after using the identity

$$b^{\frac{1}{2}} - a^{\frac{1}{2}} = (2a^{\frac{1}{2}})^{-1}((b-a) - (b^{\frac{1}{2}} - a^{\frac{1}{2}})^2))$$

one gets,

$$\nabla vhd_{1n}(\theta_0) = T_{1n} + R_{1n}$$

where

$$T_{1n} \equiv \frac{1}{4}\int_{\mathcal{R}} u_1(x|\theta_0)(f(x|\theta_0) - h_{1n}(x))dx, \tag{4.50}$$

and

$$R_{1n} \equiv \frac{1}{4}\int_{\mathcal{R}} u_1(x|\theta_0)(f^{\frac{1}{2}}(x|\theta_0) - h_{1n}^{\frac{1}{2}}(x))^2dx. \tag{4.51}$$

Similar expressions hold for $vhd_{2n}(\eta_0)$ with T_{1n} being replaced by T_{2n} and R_{1n} being replaced by R_{2n}.

Hence,

$$n^{\frac{1}{2}}(\Xi_n - \Xi_0)' = A_{1n} + A_{2n} \tag{4.52}$$

where

$$A_{1n} = n^{\frac{1}{2}}(D_{1n}^{-1}(\boldsymbol{\theta}_n^*)T_{1n}, \quad D_{2n}^{-1}(\boldsymbol{\eta}_n^*)T_{2n}) \tag{4.53}$$

$$\text{and} \quad A_{2n} = n^{\frac{1}{2}}(D_{1n}^{-1}(\boldsymbol{\theta}_n^*)R_{1n}, \quad D_{2n}^{-1}(\boldsymbol{\eta}_n^*)R_{2n}). \tag{4.54}$$

We will show that as $n \to \infty$, under further model assumptions, that (1) $A_{2n} \xrightarrow{P} \mathbf{0}$ and (2)$A_{1n} \xrightarrow{d} N_2(\mathbf{0}, \Sigma)$. We begin with the model assumptions:

(M1) The functions $u_1(\theta)s_1(\theta)$ and $u_2(\eta)s_2(\eta)$ are continuous and bounded in $L_2(.)$ at θ_0 and η_0 respectively.

(M2) The functions $\dot{u}_1\ (\theta)s_1(\theta)$ and $\dot{u}_2\ (\eta)s_2(\eta)$ are continuous and bounded in $L_2(.)$ at θ_0 and η_0 respectively.

(M3) The functions $u_1(\theta)u_1'(\theta)s_1(\theta)$ and $u_2(\eta)s_2(\eta)$ are continuous and bounded in $L_2(.)$ at θ_0 and η_0 respectively.

(M4) Let $\{\alpha_n,\ n \geq 1\}$ be a sequence diverging to infinity. Assume that

$$\lim_{n\to\infty} n \sup_{t\in \text{supp}(K)} P_{\theta_0}(|X - c_n t| > \alpha_n) = 0$$

$$\lim_{n\to\infty} n \sup_{t\in \text{supp}(K)} P_{\eta_0}(|X - c_n t| > \alpha_n) = 0$$

where $\text{supp}(K)$ is the support of the kernel density $K(\cdot)$ and X is a generic random variable with density $f(.|\theta_0)$ or $g(.|\eta_0)$ depending on where it is referenced.

(M5) Let

$$M_n(1) = \sup_{|x|\leq\alpha_n} \sup_{t\in \text{supp}(K)} \left| f^{-1}(x|\theta_0) f(x + tc_n|\theta_0) \right|$$

$$M_n(2) = \sup_{|x|\leq\alpha_n} \sup_{t\in \text{supp}(K)} \left| g^{-1}(x|\eta_0) g(x + tc_n|\eta_0) \right|.$$

Assume

$$\sup_{n\geq 1} M_n(i) < \infty \text{ for } i = 1, 2.$$

(M6) The score functions have a regular central behavior relative to the smoothing constants:i.e.

$$\lim_{n\to\infty} (n^{\frac{1}{2}} c_n)^{-1} \int_{-\alpha_n}^{\alpha_n} u_1(x|\theta_0) dx = 0 \ \text{ and } \lim_{n\to\infty} (n^{\frac{1}{2}} c_n)^{-1} \int_{-\alpha_n}^{\alpha_n} u_2(x|\eta_0) dx = 0. \tag{4.55}$$

Furthermore,

$$\lim_{n\to\infty} (n^{\frac{1}{2}} c_n^4) \int_{-\alpha_n}^{\alpha_n} u_1(x|\theta_0) dx = 0 \ \text{ and } \lim_{n\to\infty} (n^{\frac{1}{2}} c_n^4) \int_{-\alpha_n}^{\alpha_n} u_2(x|\eta_0) dx = 0. \tag{4.56}$$

(M7) The density functions are smooth in an L_2 sense; i.e.

$$\lim_{n\to\infty} \sup_{t\in Supp(K)} \int_{\mathcal{R}} (u_1(x + c_n t|\theta_0) - u_1(x|\theta_0))^2 f(x|\theta_0) dx = 0$$

and

$$\lim_{n\to\infty} \sup_{t\in Supp(K)} \int_{\mathcal{R}} (u_2(x + c_n t|\theta_0) - u_2(x|\theta_0))^2 g(x|\eta_0) dx = 0.$$

Theorem 4.1 Assume that conditions **(A0)-(A2)**,**(K1)-(K2)**, **(M1)-(M7)** hold. Then, as $n \to \infty$

$$\sqrt{n}(\Xi_n - \Xi_0) \xrightarrow{d} N_2(\mathbf{0}, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} \pi_1^{-1} I_1^{-1}(\theta_0) & 0 \\ 0 & \pi_2^{-1} I_2^{-1}(\eta_0) \end{pmatrix}. \tag{4.57}$$

where

$$I_1^{-1}(\boldsymbol{\theta}_0) = [\int (\dot{u}_1\ (x|\theta_0))(\dot{u}_1\ (x|\theta_0))' dx], \tag{4.58}$$

$$I_2^{-1}(\boldsymbol{\eta}_0) = [\int (\dot{u}_2\ (x|\eta_0))(\dot{u}_2\ (x|\eta_0))' dx] \tag{4.59}$$

where π_1 is as in (1.1) and $I_1(\theta)$ and $I_2(\eta)$ are the Fisher's information for the families $\{f(.|\theta)|\theta \in \Theta_1\}$and $\{g(.|\eta)|\eta \in \Theta_2\}$.

Proof: The proof of the Theorem will be sub-divided into various lemmas. Our first lemma studies the behavior of $D_{1n}(\theta_n^*)$ as $n \to \infty$.

Lemma 4.2 Under **(A2)-(A3)** and **(M1)-(M3)** the following hold:

1. $\lim_{n\to\infty} D_{1n}(\boldsymbol{\theta}_n^*) = 4^{-1} I_1(\boldsymbol{\theta}_0)$;

2. $\lim_{n\to\infty} D_{2n}(\boldsymbol{\eta}_n^*) = 4^{-1} I_2(\boldsymbol{\eta}_0)$.

Proof: We will only prove (i). The main idea of the proof is to show that the first term on the RHS of (4.46) converges to 0 and the second term converges to the appropriate limit. Note that the convergence of the integrands follows from the asuumed continuity of the densities and convergence of the estimates of the densities. We will show that our regularity assumptions yield convergence of the integrals themselves. Note that the first term on the RHS of (4.46) can be expressed as

$$\frac{1}{2}\int_{\mathcal{R}} (\dot{u}_1\ (x|\theta_n^*) s_1(x|\theta_n^*) - \dot{u}_1\ (x|\theta_0) s_1(\theta_0))(h_{1n}^{\frac{1}{2}}(x) - s_1(x|\theta_0) dx.$$

The above term converges to 0 by Cauchy Schwarz inequality, Theorem 3.3, assumption **(M2)** and Remark 3.2. The convergence of the second term follows along the same lines now using using assumption **(M3)** instead of **(M2)** and using the definition of the Fisher information matrix. ■

Our next lemma shows that $A_{2n} \xrightarrow{p} 0$ as $n \to \infty$ under further regularity conditions.

Lemma 4.3 Under the conditions of the theorem, $\lim_{n\to\infty} A_{2n} \xrightarrow{p} 0$.

Proof: Using Lemma 4.2, it follows that it is enough to show that R_{1n} and R_{2n} converges to 0 in probability. We will only show that R_{1n} converges to 0 in probability. We will suppress the constant and making use of an abuse of notation denote by R_{1n} the terms involving the integrals. Let us set

$$d_{n1}(x) = (s_1(x) - h_{1n}^{\frac{1}{2}}(x)). \tag{4.60}$$

Then recall that,

$$\begin{aligned} n^{\frac{-1}{2}}|R_{1n}| &\leq \frac{1}{2}\int_{\mathcal{R}} |u_1(x|\theta_0)| d_{n1}^2 dx && (4.61)\\ &\leq \frac{1}{2}\int_{-\alpha_n}^{\alpha_n} |u_1(x|\theta_0)| d_{n1}^2(x) dx + \frac{1}{2}\int_{|x|\geq\alpha_n} |u_1(x|\theta_0)| d_{n1}^2(x) dx && (4.62)\\ &= T_{1n} + T_{2n} && (4.63) \end{aligned}$$

We will now show that each of the terms are $\mathbf{o}_p(n^{\frac{-1}{2}})$ as $n \to \infty$. We begin with T_{1n}. The following estimate is needed for our analysis.

$$d_{n1}^2(x) \leq 2\{(f(x|\theta_0) - E(h_{1n}(x))^2 + (E(h_{1n}(x))^2 - h_{1n}(x))^2\}f^{-1}(x|\theta_0) \tag{4.64}$$

Using this estimate, we can write

$$T_{1n} \leq Tr_{1n}(1) + Tr_{1n}(2) \tag{4.65}$$

where

$$Tr_{1n}(1) = \int_{-\alpha_n}^{\alpha_n} |u_1(x|\theta_0)|(E(h_{1n}(x)) - h_{1n}(x))^2 f^{-1}(x|\theta_0)dx \tag{4.66}$$

and

$$Tr_{1n}(2) = \int_{-\alpha_n}^{\alpha_n} |u_1(x|\theta_0)|(f(x|\theta_0) - E(h_{1n}(x))^2 f^{-1}(x|\theta_0)dx. \tag{4.67}$$

We begin with $Tr_{1n}(1)$. Let $\epsilon > 0$ be arbitray but fixed. Then,

$$\begin{aligned} P(n^{\frac{1}{2}}Tr_{1n}(1) > \epsilon) &\leq \epsilon^{-1}n^{\frac{1}{2}}E(Tr_{1n}(1)) & (4.68)\\ &\leq \epsilon^{-1}n^{\frac{1}{2}}\int_{-\alpha_n}^{\alpha_n} |u_1(x|\theta_0)|(Var(h_{1n}(x))f^{-1}(x|\theta_0)dx & (4.69)\end{aligned}$$

Now, using **(A0)**

$$\begin{aligned} Var(h_{1n}(x)) &= E(Var(h_{n1}(x)|\mathcal{F}_n)) & (4.70)\\ &\leq \frac{1}{nc_n}E(\frac{n}{N_1(n)})\int_{\mathcal{R}} K^2(t)f(x - tc_n|\theta_0)dt. & (4.71)\end{aligned}$$

Now plugging in (4.71) in (4.66) and interchanging the order of integration (using Tonelli's Theorem)

$$P(n^{\frac{1}{2}}Tr_{1n}(1) > \epsilon) \leq C(n^{\frac{1}{2}}c_n)^{-1}M_n(1)\int_{-\alpha_n}^{\alpha_n} |u_1(x|\theta_0)|dx \tag{4.72}$$

where C is a universal constant. The result now follws from the conditions **(M6)- (M7)**. We now deal with $Tr_{1n}(2)$. To this end, we need to evaluate $(E(h_{1n}(x)) - f(x|\theta_0)^2$. Using a change of variables,two-step Taylor approximation, and **(K1)** we get

$$\begin{aligned} (E(h_{1n}(x)) - f(x|\theta_0)) &= \int_{\mathcal{R}} K(t)(f(x - tc_n|\theta_0) - f(x|\theta_0))dt & (4.73)\\ &= \int_{\mathcal{R}} K(t)\frac{(tc_n)^2}{2}f''(x_n^*(t)|\theta_0)dt. & (4.74)\end{aligned}$$

Now plugging in (4.74) into (4.67) and using conditions **(M3)**, and **(M7)** one gets

$$n^{\frac{1}{2}}Tr_{1n}(2) \leq Cn^{\frac{1}{2}}c_n^4\int_{-\alpha_n}^{\alpha_n} |u_1(x|\theta_0)|dx \tag{4.75}$$

where C is a universal constant. Convergence of (4.75) to 0 now follows from condition **(M6)**. We next deal with T_{2n}. To this end, observe that

$$n^{\frac{1}{2}}T_{2n} = \int_{|x|\geq\alpha_n} |u_1(x|\theta_0)|(f(x|\theta_0) + h_{1n}(x) + s_1(x|\theta_0)h_{1n}^{\frac{1}{2}}(x))dx. \tag{4.76}$$

We will show that the RHS of the above equation converges to 0 as $n \to \infty$. We begin with the first term. Note that, by Caucy-Schwarz inequality

$$n(\int_{|x|\geq\alpha_n} |u_1(x|\theta_0)f(x|\theta_0)dx)^2 \leq (\int_{|x|\geq\alpha_n} u_1^2(x|\theta_0)f(x|\theta_0)dx)\{nP_{\theta_0}(|X| \geq \alpha_n)\} \tag{4.77}$$

which converges to 0 by **(M4)**. As for the second term, note that, a.s.

$$(\int_{|x|\geq\alpha_n} |u_1(x|\theta_0)|h_{1n}(x)dx)^2 \leq \int_{|x|\geq\alpha_n} u_1^2(x|\theta_0)|h_{1n}(x)dx.$$

Now taking the expectation and using Cauchy-Schwarz inequality, one can show that

$$nE(\int_{|x|\geq\alpha_n} |u_1(x|\theta_0)h_{1n}(x)dx)^2 \leq (\int_{\mathcal{R}} K(t) \int_{\mathcal{R}} u_1^2(x - c_n t)f(x|\theta_0)dxdt)m_n \tag{4.78}$$

where $m_n \equiv n \sup_{z\in Supp(K)} P_{\theta_0}(|X - c_n z| \geq \alpha_n)$. The convergence to 0 of (4.78) now follows from conditions **(M4)**. Finally, by yet another application of the Cauchy-Schwarz inequality,

$$nE(\int_{|x|\geq\alpha_n} |u_1(x|\theta_0)|h_{1n}^{\frac{1}{2}}(x)s_1(x|\theta_0)dx) \leq (\int_{\mathcal{R}} u_1^2(x|\theta_0)f(x|\theta_0)dx)m_n. \tag{4.79}$$

Convergence of (4.79) to 0 follows from **(M4)**. This completes the proof that $R_{1n} \to 0$ as $n \to \infty$. ∎

Our next lemma studies the asymptotic limit behavior of $n^{\frac{1}{2}}(T_{1n}, T_{2n})$ defined in (4.50). The idea of proof is to first show that the limit distribution of $(n^{\frac{1}{2}}T_{1n}, n^{\frac{1}{2}}T_{2n})$ does not depend on c_n and then use the Cramer-Wold device to obtain the limiting joint distribution. The main difficulty is of course, the terms T_{1n} and T_{2n} are dependent due to the fact that the sample sizes are dependent random variables. We address this difficulty using Kolmogrov's maximal inequality. We begin with a lemma that shows that "in a second order limiting sense" we can approximate the distribution of $n^{\frac{1}{2}}(T_{1n}, T_{2n})$ with terms involving the empirical distribution. More precisely,

Lemma 4.4 Under **(M7)** the following holds:

$$\lim_{n\to\infty} E(4n^{\frac{1}{2}}T_{1n} - \frac{n^{\frac{1}{2}}}{N_1(n)} \sum_{j\in\mathcal{A}_n(1)} u_1(X_{1,\nu(1,j)}|\theta_0))^2 = 0. \tag{4.80}$$

$$\lim_{n\to\infty} E(4n^{\frac{1}{2}}T_{2n} - \frac{n^{\frac{1}{2}}}{N_2(n)} \sum_{j\in\mathcal{A}_n(2)} u_2(X_{2,\nu(1,j)}|\eta_0))^2 = 0. \tag{4.81}$$

Proof: One can show that,

$$n^{\frac{1}{2}}(4T_{1n} - \frac{1}{N_1(n)} \sum_{j \in \mathcal{A}_n(1)} u_1(X_{1,\nu(1,j)}|\theta_0)) = n^{\frac{1}{2}}(\sum_{j \in \mathcal{A}_n(1)} \int_{\mathcal{R}} u_{1n}(X_{1\nu(1j)}, t)K(t)dt) \tag{4.82}$$

where

$$u_{1n}(x,t) = (((u_1(x + tc_n|\theta_0) - u_1(x|\theta_0))). \tag{4.83}$$

The result is proved using Cauchy-Schwarz inequality, Minkowski's inequlaity (along the lines of Proof of Lemma 4.3), and the bounded convergence theorem using **(M7)**. ∎

Our next lemma studies the asymptotic limit distribution of $\sqrt{n}(S^1_{N_1}, \ \ S^2_{N_2})$ where

$$S^1_{N_1} = \frac{1}{N_1(n)} \sum_{j \in \mathcal{A}_n(1)} u_1(X_{1\nu(ij)}|\theta_0). \tag{4.84}$$

and $S^2_{N_2}$ is defined similarly.

Lemma 4.5 The normalized scores converge in distribution to a bivariate normal distribution, i.e.

$$\lim_{n \to \infty} \sqrt{n}(S^1_{N_1}, \ \ S^2_{N_2}) \stackrel{d}{\Rightarrow} N_2(\mathbf{0}, \Sigma) \tag{4.85}$$

where

$$\Sigma = \begin{pmatrix} \pi_1 I_1^{-1}(\theta_0) & 0 \\ 0 & \pi_2 I_2^{-1}(\eta_0) \end{pmatrix}. \tag{4.86}$$

Proof: We will use the Cramer-Wold device. Let l_1 and l_2 be $1 \times p$ vectors of constants. The linear combination of $S^1_{N_1}$ and $S^2_{N_2}$ is then

$$\sqrt{n}(\sum_{i=1}^{2} \frac{1}{N_i} \sum_{j=1}^{N_i} l_i u_i(X_{i\nu(ij)})) \tag{4.87}$$

Thus to complete the proof we need to show that the term

$$\sqrt{n} \sum_{i=1}^{2} \frac{1}{N_i} \sum_{j \in \mathcal{A}_n(i)} l_i u_i(X_{i\nu(ij)})$$

converges a normal distribution. Now

$$\sqrt{n} \sum_{i=1}^{2} \frac{1}{N_i} \sum_{j \in \mathcal{A}_n(i)} l_i u_i(X_{i\nu(ij)}) \ \ = \ \ G_{n,1} + G_{n,2} + G_{n,3} + G_{n,4} \tag{4.88}$$

where

$$G_{n,1} \ \ = \ \ \frac{1}{N_1} \sum_{j \in \mathcal{A}_n(1)} l_1' u_1(X_{1\nu(ij)}) = \frac{1}{N_1} \sum_{j=1}^{[n \cdot \pi_1]} l_1' u_1(X_{1\nu(ij)})$$

$$
\begin{aligned}
G_{n,2} &= \frac{1}{N_2}\sum_{j=1}^{[n\pi_2]} l_2' u_2(X_{1\nu(2j)}) \\
G_{n,3} &= \frac{1}{N_1}\sum_{[\pi_1 n]\wedge N_1 \le j \le N_1 \vee [\pi_1 n]} l_i' u_1(X_{1\nu(1j)}) \\
G_{n,4} &= \frac{1}{N_2}\sum_{[\pi_2 n]\wedge N_2 \le j \le N_2 \vee [\pi_2 n]} l_2' u_2(X_{2\nu(2j)})
\end{aligned}
$$

By Melfi's results and central limit theorem for i.i.d. random variables $G_{n,1} + G_{n,2}$ converges to a linear combination of Gaussian random vectors. We will now show that $G_{n,3} + G_{n,4}$ converges to 0 in probability. Now, for any $\delta > 0$,

$$
\begin{aligned}
P(|G_{n,3}| > \epsilon) &= P(|G_{n,3}| > \epsilon : |\frac{N_1}{n} - \pi_1| < \delta) + P(|G_{n,3}| > \epsilon : |\frac{N_1}{n} - \pi_1| > \delta) \\
&= \text{(i)} \qquad\qquad\qquad\qquad + \text{(ii)}
\end{aligned}
$$

Let $E(n,\delta) = \{\omega : |\frac{N_1(w)}{n} - \pi_1| < \delta\}$.
Note that

$$
\begin{aligned}
\text{(i)} &= P\left\{\left(\frac{1}{N_1}\sum_{j=N_1\wedge[n\pi_1]}^{N_1\vee[n\pi_1]} \psi_1(X_{1\nu(1j)})\right) > \epsilon : \quad E(n,\delta)\right\} \\
&\le P\left(\frac{1}{n(\pi_1+\delta)}\sum_{j=[n(\pi_1-\delta)]}^{n(\pi_1+\delta)} \psi_1(X_{1\nu(1j)}) > \epsilon\right) \\
&\le P\left(\max_{1\le k\le n(\pi_1+\delta)}\sum_{j=[n(\pi_1-\delta)]}^{k} \psi_1(X_{1\nu(1j)}) \ge n(\pi_1+\delta)\epsilon\right) \\
&\le \frac{C}{n^2} \quad \to 0 \ \text{ as } \ n\to\infty. \qquad (4.89)
\end{aligned}
$$

where (4.89) follows from Kolmogorov's maximal inequality (Chung (1974) P. 116.) Similar argument shows that $G_{n,4} \xrightarrow{P} 0$. As for (ii)

$$
\text{(ii)} \le P(|\frac{N_1}{n} - \pi_1| > \delta) \to 0 \ \text{ as } \ n\to\infty.
$$

■

Proof of the Theorem. From (4.52) , (4.53), and (4.50), it is enough to establish the asymptotic normality of A_{1n}. To this end, note that

$$
A_{1n} = Diag(D_{1n}^{-1}(\theta_n^*), D_{2n}^{-1}(\eta_n^*))(T_{1n}, T_{2n})'. \qquad (4.90)
$$

Hence, using Lemma 4.2, Lemma 4.4 and Lemma 4.5 and a straight forward calculation, the Theorem follows. ■

We next move on to discuss robustness.

5 Robustness of MHDE

In this section we deal with the robustness of the MHDE. We describe the robustness properties through a study of the α-influence function and the breakdown point. We begin with the α-influence function. We will denote by $F(\cdot|\Xi,\ \alpha,\ z)$ the contaminated model, *i.e.*

$$F(\cdot|\Xi,\ \alpha,\ z) = (1-\alpha)F(\cdot|\xi) + \alpha U_z \tag{5.91}$$

where

$$U_z = \begin{pmatrix} U_{Z_1} \\ U_{Z_2} \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

U_{Z_i} are uniform densities on the interval $(Z_i - \epsilon, Z_i + \epsilon)$ where $\epsilon > 0$. Note that $f(\cdot|\theta\ ,\ \alpha_1,\ Z_1)$ represents a $(1-\alpha_1)\%$ contamination with distant "outliers". Similarly, $g(\cdot|\eta\ ,\ \alpha_2,\ Z_2)$ represents a $(1-\alpha_2)\%$ contamination with distant "outliers". Our first main result of this section is contained in the following theorem.

Theorem 5.1. Assume that the conditions of Theorem 4.1. hold. If $T(F(\cdot|\Xi,\ \alpha,\ z)$ is unique for all z, then

(i) $T(F(\cdot|\Xi,\ \alpha,\ z)$ is a bounded continuous function of z such that

$$\lim_{z\to\infty} T(F(\cdot|\Xi,\ \alpha,\ z) = \Xi. \tag{5.92}$$

Furthermore,

(ii)

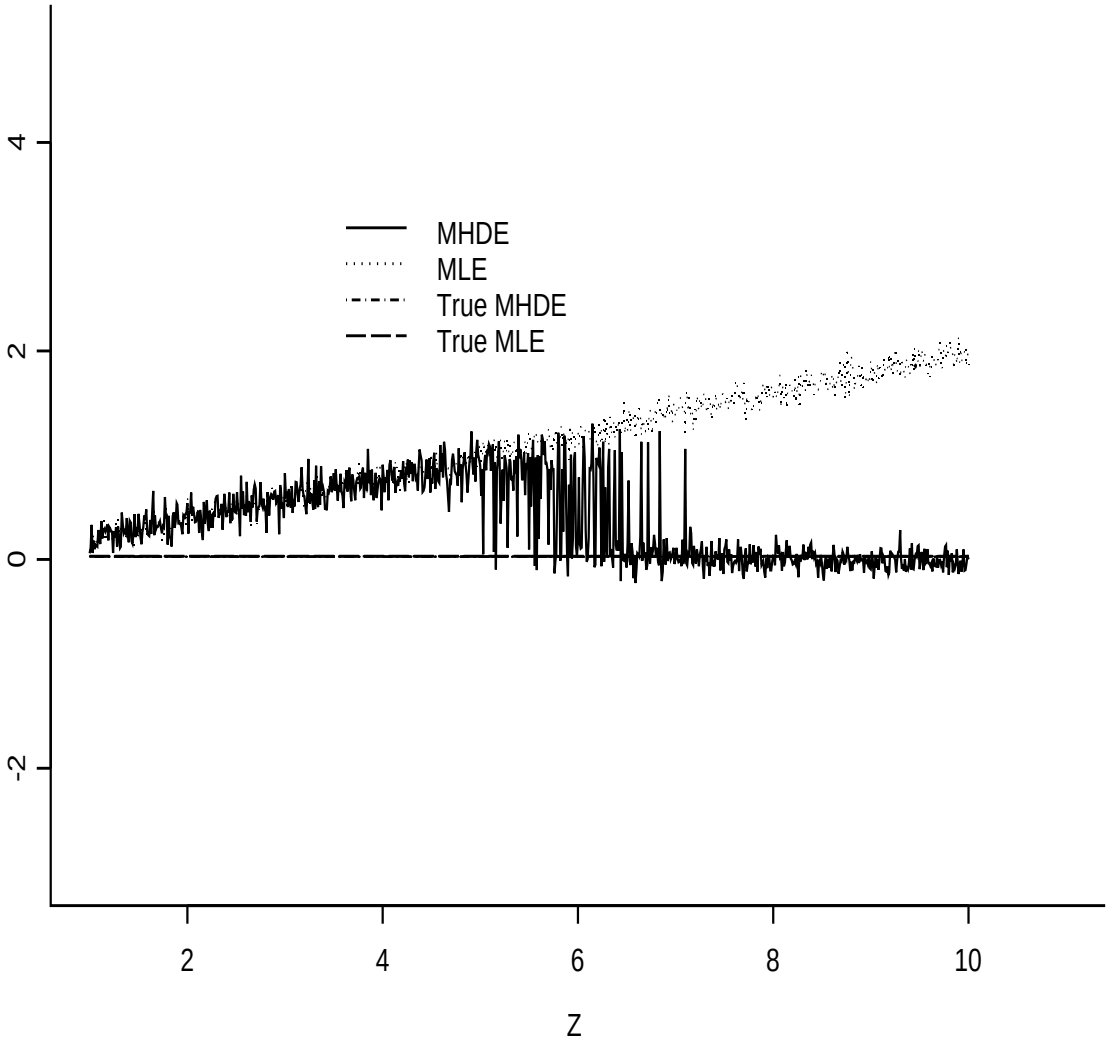
$$\lim_{\alpha\to 0}(T(F(\cdot|\Xi,\ \alpha,\ z) - \Xi)\alpha^{-1} = RF_T(z)$$

where

$$RF_T(z) = \begin{pmatrix} (I_1(\theta))^{-1}[\int_{\Re} U_{Z_1}(x)\psi_1(x|\theta)dx] \\ (I_2(\boldsymbol{\eta}))^{-1}[\int_{\Re} U_{Z_2}(x)\psi_2(x|\eta)dx] \end{pmatrix}.$$

Proof. The proof is a straight forward modification of Beran's proof applied component-wise with adjustments to non-compact parameter spaces and hence is omitted.

Remark 3.8.2. The functional T viewed as a function of z is called the α-influence curve.



Graph 5.1: The following graph represents the α-influence curve with 20% contamination.

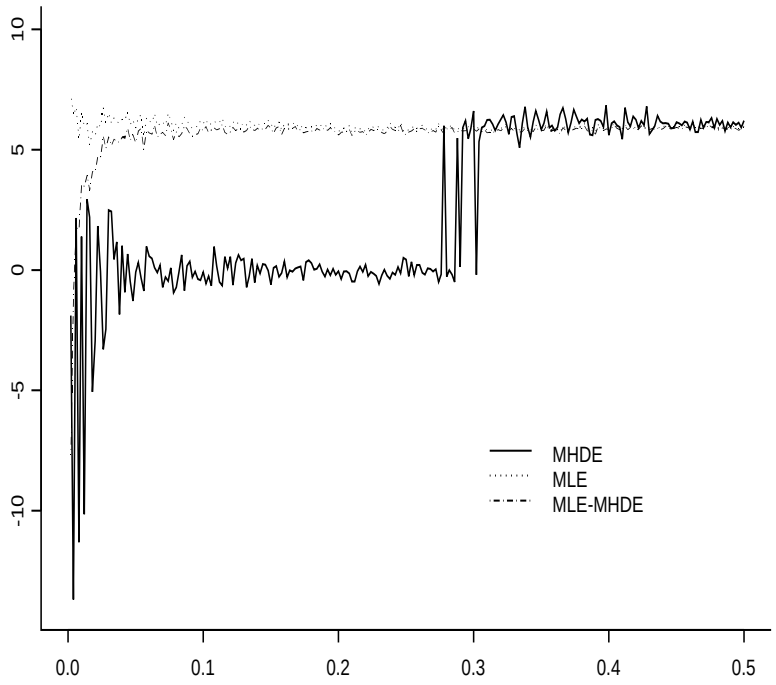
Remark 5.2. Note that $F(\cdot|\Xi, \alpha,\ z)$ models an experiment where the observations are mixed with approximately $\alpha\%$ gross errors located near z. The above theorem compares $T(F(\cdot|\Xi, \alpha,\ z))$ with $T(F(\cdot|\Xi)) = \Xi$.

Remark 5.3. The Graph 5.1 on the last page described the influence of z for the Hellinger distance estimator. Note also that the graphs for various α's change dramatically, implying that the convergence of the α-influence curve need not be uniform in z. To contrast our results with the MLE, we note that the α-influence curve of the MLE is unbounded since

$$|\hat{\theta}_{MLE,z}| \to \infty$$

as $z \to \infty$. This can be alsobe seen from the Graph 5.2.

We now move on to describe the breakdown point of MHDE. One can show, in a manner analogous to Simpson(1987) that the asymptotic breakdown point of the MHDE is 1/2. Since the ideas involve no new novelties, we refer the reader to Simpson's article for deatils. We satisfy by presenting the following graph that shows a relative change in the estimator due to the change in the contamination proportion.



Graph 5.2: The above graph represents the relative change in the estimator due to the change in the contamination proportion. α is the percentage of contamination.

6 Computational Algorithm

In this section we introduce a numerical method to solve equations (4.41) for θ and η. Assuming $f(x|\theta)$ to be the Normal distribution with mean μ and variance σ^2,Beran (1977) applied the Newton-Raphson method for solving MHDE. Unlike the Newton-Raphson method algorithm, which requires calculation of the second-order derivatives, one step Monte Carlo approximation method is easier to implement. The method is motivated as follows: Recall that finding the MHDE of θ is equivalent to finding the θ that maximizes the following:

$$\int (f(x|\theta)^{1/2}(h_{in_i}(x))^{1/2}dx = \int \{\frac{(f(x|\theta)^{1/2}}{(h_{in_i}(x))^{1/2}}\}(h_{in_i}(x))dx.$$

By strong law of large numbers the above integral can be approximated by

$$\frac{1}{M}\sum_{j=1}^{M}(\frac{f_\theta(y_{ij})}{h_{in_i}(y_{ij})})^{1/2}, \tag{6.93}$$

where M is the number of the Monte Carlo samples and $y_{ij} \sim h_{in_i}$. We need to find the value of θ that maximizes (6.93). When the underlying distribution of f_θ is $N(\mu, \sigma^2)$, (6.93) becomes the following :

$$\frac{1}{M}\sum_{j=1}^{M}\frac{w_{ij}}{\sqrt[4]{2\pi\sigma^2}}\exp(-\frac{1}{4\sigma^2}(y_{ij}-\mu)^2), \quad w_{ij} = \frac{1}{\sqrt{h_{in_i}(y_{ij})}}. \tag{6.94}$$

Taking the partial derivative of (6.94) with respect to μ and σ^2 and setting them to 0, we obtain the following recursive equations for μ and σ^2, *viz.*,

$$\hat{\mu}_{(m+1)} = \frac{\sum_{j=1}^{M} w_{ij}\exp(-\frac{1}{4\hat{\sigma}^2_{(m)}}(y_{ij}-\hat{\mu}_{(m)})^2)y_{ij}}{\sum_{j=1}^{M} w_{ij}\exp(-\frac{1}{4\hat{\sigma}^2_{(m)}}(y_{ij}-\hat{\mu}_{(m)})^2)} \tag{6.95}$$

and

$$\hat{\sigma}^2_{(m+1)} = \frac{\sum_{j=1}^{M} w_{ij}\exp(-\frac{1}{4\hat{\sigma}^2_{(m)}}(y_{ij}-\hat{\mu}_{(m)})^2)(y_{ij}-\hat{\mu}_{(m)})^2}{\sum_{j=1}^{M} w_j\exp(-\frac{1}{4\hat{\sigma}^2_{(m)}}(y_{ij}-\hat{\mu}_{(m)})^2)}. \tag{6.96}$$

If the kernel K is a standard normal density, we have

$$\begin{aligned}
h_{in_i}(x) &= \frac{1}{n_i c_n}\sum_{z=1}^{n_i} K\left\{\frac{x - X_{i\nu(iz)}}{c_n}\right\} \\
&= \frac{1}{n_i c_n}\sum_{z=1}^{n_i}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{x-X_{i\nu(iz)}}{c_n})^2) \\
&= \frac{1}{n_i}\sum_{z=1}^{n_i}\frac{1}{\sqrt{2\pi c_n^2}}\exp(-\frac{1}{2}(\frac{x-X_{i\nu(iz)}}{c_n})^2) \\
&= \frac{1}{n_i}\sum_{z=1}^{n_i}\phi(X_{i\nu(iz)}, c_n^2),
\end{aligned}$$

where ϕ is the normal density with mean equal to $X_{i,\nu(iz)}$ and variance equal to c_n^2. Thus, $h_{in_i}(x)$ is a mixture of normal densities with mixing proportion $\frac{1}{n_i}$. Therefore, in the first step of the algorithm, we generate a random variable y_{ij} which has the distribution $h_{in_i}(x)$.

Note that in the update formulas (19) and (20), $w_{ij} = \frac{1}{\sqrt{h_{i,n_i}(y_{ij})}}$ which depends on the choice of the kernel density K. When K is chosen to be standard normal density, we have

$$w_{ij} = \left[\frac{1}{n_i c_n}\sum_{z=1}^{ni}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{y_{ij}-X_{i\nu(iz)}}{c_n})^2)\right]^{-1/2}.$$

If $K(\cdot)$ were Epanechnikov kernel the weight reduces to

$$w_{ij} = \left[\frac{1}{n_i c_n}\sum_{z=1}^{ni}0.75(1-(\frac{y_{ij}-X_{i\nu(iz)}}{c_n})^2)\right]^{-1/2}, \quad \left|\frac{y_{ij}-X_{i\nu(iz)}}{c_n}\right| < 1.$$

Let us describe how to generate the Epanechnikov random variable. From Devroye (1986 P.72), we know that if U_1, U_2, U_3 and U_4 are i.i.d. uniform$[0,1]$ random variables, then for $a > 1$, $U_1^{\frac{1}{a}}U_2$ has density

$$\frac{a}{a-1}(1-x^{a-1})I_{[0<x<1]};$$

also, $(-U_3^{\frac{1}{a}})U_4$ has density

$$\frac{a}{a-1}(1-x^{a-1})I_{[-1<x<0]}$$

In our case, letting $a = 3$, and randomizing between $U_1^{\frac{1}{a}}U_2$ or $(-U_3^{\frac{1}{a}})U_4$ with equal probability we obtain the Epanechnikov random variable with mean 0 and variance 1, denoted as Epa(0,1). Therefore, Epanechnikov random variable with mean $X_{i,\nu(i,z)}$ and variance c_n^2 can be obtained by Epa(0,1)*c_n+$X_{i,\nu(i,z)}$. The one step Monte-Carlo approximation algorithm can be described as follows:

1. Generate random variables for each data point from the kernel density with mean $X_{i\nu(iz)}$ and variance c_n^2. Choose one of then with equal probability ($\frac{1}{n_i}$) and retain it. Repeat M times. Using the initial values for μ and σ, *viz.*, $\hat{\mu}^{(0)}$ =median$\{X_{i\nu(iz)}\}$ and $\hat{\sigma}^{(0)} = (0.674^{-1})$median$\{|X_{i\nu(iz)} - \hat{\mu}^{(0)}|\}$.

2. Obtain the updates using (19) and (20).

3. When $|\hat{\mu}_{(m+1)} - \hat{\mu}_{(m)}| < \epsilon$ and $|\hat{\sigma}_{(m+1)} - \hat{\sigma}_{(m)}| < \epsilon$ for small ϵ, say 10^{-6} then stop; else go to step2.

7 Simulation Results

In this section, we will present simulation results which were carried out using SAS software.The simulations compares the efficiency between MHDE and MLE with outlier, and

incorporating the randomized play the winner design with two treatments. We start with an urn containing 5 ball of each type and assume treatment A has success probability p_1 and treatment B has success probability p_2. Let N_1^0 denote the number of type A balls in the urn at the beginning of the trial and N_2^0 denotes the number of type B balls in the urn at the beginning of the trial. Let N_1^i and N_2^i denote the number of type A and B ball, after the i^{th} patient's response is observed and the urn has been updated. The simulation procedure works as follows:

1. Generate a uniform(0,1) random variable, say u_1.

2. If $u_1 > \frac{N_1^i}{(N_1^i+N_2^i)}$, assign patient $i+1$th to treatment B and generate a N(5,3) random variable, representing the secondary variable. Otherwise, assign patient to treatment A and generate a N(0,1) random variable, representing the secondary variable.

3. Generate a uniform(0,1) random variable, say u_2. If the treatment assignment in step 2 is A and $u_2 < P_1$, then call this treatment a success, and add one type A ball to the urn. Otherwise, add a type B ball. If the treatment assignment in step 2 is B, we will update the urn similarly.

4. Repeat steps 1, 2 and 3 for 30 times to represent a sample of size 30.

5. Calculate MHDE and MLE for both treatments.

6. Repeat the above steps 1000 times.

In our simulation results we used Epanechnikov kernel to obtain the MHDE. The window width c_n was chosen by first fitting different values of c_n for different sample sizes (around the average smaple size). Note that the average sample size changes depending on the the sample size n and the values of design parameters. After a best fitting c_n was determined, it was held fixed in the simulations for those choices of the design parameters.

Our simulations illustrate the probability that the MHDE and MLE will fall into the true 95% confidence interval for both the treatments A and B. Note that the true 95% confidence interval for treatment A is given by $(0 - 1.96\frac{1}{\sqrt{N_1}}, 0 + 1.96\frac{1}{\sqrt{N_1}})$, where N_1 is the number of patients allocated to treatment A. To illustrate the robustness property of the MHDE, we will change some of the treatment A's responses to come from a normal distributions with means ranging between 2 and 6. The design parameters p_1 and p_2, the probability of sucess on treatments A and B play an important role through the behavior of the function $Q = p_1 + p_2 - \frac{3}{2}$. The case when $Q < 0$ is a standard case, in the sense that there is a central limit behavior for the functional $\frac{N_1}{n}$ with normalization $\sqrt{n}$. The cases $Q = 0$ and $Q > 0$ lead to non-stanadard results in that either the normalization is different or the limit is a non-normal distribution ([.]) and our simulations study if there is any impact of the design in the probability that

the estimator belongs to the confidence interval. Note that even though the limit results do not depend the design parameters, (due to the fact that the problem is reduced to a random sample size problem and the fact that the random sample size diverges to infinity is all that matters for the limit theory to hold) "true conditional confidence intervals" depend on the design through the random sample sizes. Table 7.1 and Table 7.2 below compare the behavior of MHDE and MLE when $Q < 0$.

Table 7.1: Results for the RPWD. The probability that the MHDE and MLE fall in the true confidence intervals with outlier from treatment A. The number of outliers equal 1, 2 and 3 with data from N(2,1) to N(6,1). Significant level=0.05, $N_1(0) = N_2(0) = 5$, $\alpha = 1$, $n = 30$, 1000 simulations and $p_1 = 0.50,\ p_2 = 0.50$

outliers	Estimator	N(2,1)		N(3,1)		N(4,1)		N(5,1)		N(6,1)		N(7,1)	
		A	B	A	B	A	B	A	B	A	B	A	B
1	MHDE	0.89	0.93	**0.89**	0.95	**0.90**	0.93	**0.91**	0.94	**0.92**	0.94	**0.94**	0.94
	MLE	0.92	0.94	**0.88**	0.95	**0.82**	0.94	**0.73**	0.94	**0.66**	0.94	**0.54**	0.95
2	MHDE	0.79	0.94	**0.75**	0.93	**0.74**	0.95	**0.81**	0.94	**0.88**	0.93	**0.90**	0.94
	MLE	0.79	0.95	**0.67**	0.94	**0.45**	0.95	**0.29**	0.96	**0.14**	0.93	**0.07**	0.95
3	MHDE	**0.67**	0.92	**0.48**	0.94	**0.39**	0.94	**0.52**	0.95	**0.62**	0.94	**0.77**	0.94
	MLE	**0.64**	0.93	**0.37**	0.94	**0.13**	0.94	**0.04**	0.96	**0**	0.95	**0**	0.96

The numbers in the bold font represent the cases that the proportion of times that MHDE falls into true CI is higher than MLE. From Table 7.1, we see that as the values of the outliers become larger, the probability of MLE falling into the true confidence interval is much smaller than MHDE. Increasing the number of outliers makes the situation even worse.

Table 7.2: Results for the RPWD. The probability that the MHDE and MLE fall in the true confidence intervals with outlier from treatment A. The number of outliers equal 1, 2 and 3 with data from N(2,1) to N(6,1). Significant level=0.05, $N_1(0) = N_2(0) = 5$, $\alpha = 1$, $n = 30$, 1000 simulations and $p_1 = 0.80,\ p_2 = 0.20$

outliers	Estimator	N(2,1)		N(3,1)		N(4,1)		N(5,1)		N(6,1)		N(7,1)	
		A	B	A	B	A	B	A	B	A	B	A	B
1	MHDE	0.94	**0.94**	0.91	0.95	**0.89**	0.93	**0.92**	0.93	**0.93**	0.93	**0.94**	0.93
	MLE	0.95	**0.93**	0.91	0.95	**0.85**	0.95	**0.79**	0.95	**0.74**	0.94	**0.63**	0.94
2	MHDE	0.88	0.93	**0.84**	0.95	**0.82**	0.94	**0.84**	0.94	**0.91**	0.92	**0.92**	0.92
	MLE	0.88	0.95	**0.77**	0.95	**0.57**	0.95	**0.39**	0.96	**0.21**	0.94	**0.12**	0.95
3	MHDE	**0.75**	0.92	**0.56**	0.93	**0.59**	0.94	**0.92**	0.93	**0.93**	0.93	**0.90**	0.93
	MLE	**0.73**	0.93	**0.42**	0.95	**0.23**	0.95	**0.79**	0.95	**0.70**	0.94	**0**	0.95

Table 7.3: Results for the RPWD. The probability that the MHDE and MLE fall in the true confidence intervals with outlier from treatment A. The number of outliers equal 1, 2 and 3 with data from N(2,1) to N(6,1). Significant level=0.05, $N_1(0) = N_2(0) = 5$, $\alpha = 1$, $n = 30$, 1000 simulations and $p_1 = 0.75,\ p_2 = 0.75$

outlier		N(2,1)		N(3,1)		N(4,1)		N(5,1)		N(6,1)	
		A	B	A	B	A	B	A	B	A	B
1	MHDE	0.898	0.926	**0.881**	0.928	**0.889**	0.933	**0.905**	0.929	**0.922**	0.928
	MLE	0.906	0.951	**0.857**	0.948	**0.796**	0.951	**0.703**	0.952	**0.608**	0.950
2	MHDE	**0.799**	0.927	**0.726**	0.925	**0.737**	0.927	**0.806**	0.930	**0.852**	0.929
	MLE	**0.791**	0.947	**0.615**	0.950	**0.421**	0.950	**0.261**	0.950	**0.135**	0.953
3	MHDE	**0.631**	0.927	**0.460**	0.931	**0.487**	0.927	**0.628**	0.928	**0.937**	0.927
	MLE	**0.605**	0.950	**0.328**	0.949	**0.137**	0.949	**0.043**	0.951	**0.011**	0.950

Instead of assuming $p_1 = p_2 = 0.5$, the simulation results in Table 7.2 above assume that $p_1 = 0.8$ and $p_2 = 0.2$, which means treatment A has a higher success probability than treatment B. From Table 7.2, we notice that the effect of outliers is similar to Table 7.1. However, due to more patients being assigned to treatment A, the results in Table 7.2 are not as dramatic as in Table 7.1.

Our next tables compares the behavior of MHDE and MLE when $Q > 0$.

Table 7.4:. Results for the RPWD. The probability that the MHDE and MLE fall in the true confidence intervals with outlier from treatment A. The number of outliers equal 1, 2 and 3 with data from N(2,1) to N(6,1)Significant level=0.05, $N_1(0) = N_2(0) = 5$, $\alpha = 1$, $n = 30$, 1000 simulations and $p_1 = 0.75,\ p_2 = 0.78$

outlier		N(2,1)		N(3,1)		N(4,1)		N(5,1)		N(6,1)	
		A	B	A	B	A	B	A	B	A	B
1	MHDE	0.895	0.925	**0.878**	0.929	**0.886**	0.927	**0.908**	0.925	**0.908**	0.927
	MLE	0.908	0.951	**0.858**	0.951	**0.779**	0.947	**0.697**	0.949	**0.600**	0.948
2	MHDE	**0.785**	0.926	**0.899**	0.928	**0.717**	0.925	**0.787**	0.924	**0.843**	0.925
	MLE	**0.777**	0.948	**0.560**	0.952	**0.407**	0.950	**0.278**	0.949	**0.126**	0.949
3	MHDE	**0.620**	0.923	**0.452**	0.923	**0.473**	0.925	**0.609**	0.930	**0.716**	0.931
	MLE	**0.601**	0.948	**0.321**	0.949	**0.132**	0.948	**0.040**	0.951	**0.010**	0.949

Table 7.5: Results for the RPWD. The probability that the MHDE and MLE fall in the true confidence intervals with outlier from treatment A. The number of outliers equal 1, 2 and 3 with data from N(2,1) to N(6,1). Significant level=0.05, $N_1(0) = N_2(0) = 5$, $\alpha = 1$, $n = 30$, 1000 simulations and $p_1 = 0.75,\ p_2 = 0.79$

outlier		N(2,1)		N(3,1)		N(4,1)		N(5,1)		N(6,1)	
		A	B	A	B	A	B	A	B	A	B
1	MHDE	0.896	0.931	**0.875**	0.933	**0.882**	0.920	**0.895**	0.929	**0.910**	0.927
	MLE	0.909	0.950	**0.851**	0.954	**0.781**	0.944	**0.692**	0.953	**0.584**	0.949
2	MHDE	**0.784**	0.924	**0.693**	0.934	**0.712**	0.933	**0.781**	0.931	**0.843**	0.933
	MLE	**0.780**	0.950	**0.584**	0.952	**0.399**	0.952	**0.235**	0.952	**0.121**	0.956
3	MHDE	**0.612**	0.929	**0.440**	0.932	**0.464**	0.926	**0.597**	0.931	**0.712**	0.925
	MLE	**0.587**	0.949	**0.308**	0.952	**0.132**	0.951	**0.043**	0.952	**0.009**	0.950

From the above simulation results we notice that there seems to be a significant drop in the coverage for both MLE and MHDE when Q is near 0 when the data are contaminated by outliers. This manifestation of low coverage persisted even when the number of simulations were increased to 5000. Further studies have indicated that there is an interaction between the design parameters and the performance of the estimators related to the parameters of the "response" variables generated using a RPWD. Methods for quantifying this interaction are being investigated by the authors.

8 Concluding Remarks

In this paper we developed Hellinger distance methodology for the analysis of data from a clinical trial conducted using randomized play the winner designs. Our theoretical and simulation results indicate that the Hellinger distance methodology is competitor to the maximum likelihood methodology and is in fact a method of choice for problems involving adaptive sampling. The extension of the methodology to general disparities and to general adaptive designs allowing for delayed response and a general patient recruitment process is considered in a separate manuscript.

Acknowledgements: Authors thank the two anonymous referees for a careful and a detailed reading of the manuscript and several critical questions and useful suggestions.

References:

1. Athreya, K.B. and Karlin, S. (1968). Embedding of urn schemes into continuous time Markov Branching processes and related limit theorems, The Annals of Mathematical Statistics, 39 1801-1817.

2. Athreya, K.B. and Vidyashankar, A.N. (1995). Large deviation rates for supercritical and critical branching processes, Classical and Modern Branching Processes, The IMA Volumes in Mathematics and its Applications, 84 1-18.

3. Basu, Ayanendranath and Lindsay, Bruce G. (1994). Minimum disparity estimation for continuous models: efficiency, distribution and robustness, Ann. Inst. Statist. Math., 46 683-705.

4. Basu, Ayanendranath and Sarkar, Sahadeb. (1994). The trade-off between robustness and efficiency and the effect of model smoothing in minimum disparity inference. J. Statist. Comput. Simul. 50 173-185.

5. Basu, Ayanendranath, Sarkar, Sahadeb and Vidyashankar, A.N.. (1997). Minimum negative exponential disparity estimation on parametric models, Journal of Statistical Planing and Inference, 58 349-370.

6. Bai, Z.D., Hu,Feifang and Rosenberger, William F.. (2002). Asymptotic properties of adaptive designs for clinical trials with delay response, The Annals of Statistics, 30 122-139.

7. Beran, Rudolf. (1977). Minimum Hellinger distance estimates for parametric models, Annals of Statistics, 5 445-463.

8. Cutler, Adele and Cordero-Braña, Oiga I. (1996). Minimum Hellinger distance estimation for finite mixture models, Journal of the American Statistical Association, 91 1716-1723.

9. Chung, K.L.(1974). A Course in Probability Theory, Academic Press.

10. Devroye, Luc. (1986). Non-Uniform Random Variate Generation, Springer-Verlag.

11. Devroye, Luc. (1987). A Course in Density Estimation, Birkhauser Boston.

12. Ivanova, Anastasia, Rosenberger, William F. Durham S. and Flournoy, Nancy.(2000). A Birth and Death Urn for randomized clinical trials: asymptotic methods, Biostatistics, 62 104-118

13. Ivanova, Anastasia and Durham S.. (2002). The Drop the Loser Rule. Submitted.

14. Kupfer, D.J. (1976). REM Latency: A Psychobiological Marker for Primary Depressive Disease, Biological Psychiatry, 11 159-174.

15. Lindsay, Bruce G. (1994). Efficiency verses robustness: the case for minimum Hellinger distance and related methods, Annals of Statistics, 22 1081-1114.

16. Markatou, Marianthi, Basu, Ayanedranath and Lindsay, Bruce. (1997). Weighted likelihood estimating equations: the discrete case with applications to logistic regression. Robust statistics and data analysis, II. J. Statist. Plann. Inference 57 215-232.

17. Melfi, Vincent F. and Page Connie. (2000). Estimation after adaptive allocation, Journal of Statistical Planing and Inference, 87 353-363

18. Rosenberger, William F., Flournoy, Nancy and Durham, Stephen D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs, Journal of Statistical Planing and Inference, 60 69-76.

19. Simpson, Douglas G. (1987). Minimum Hellinger distance estimation for the analysis of count data, Journal of the American Statistical Association, 82 802-807.

20. Simpson, Douglas G. (1989). Hellinger deviance tests: efficiency breakdown points, and examples,Journal of the American Statistical Association, 84 107-113.

21. Sriram, T. N. and Vidyashankar, A. N. (2000). Minimum Hellinger distance estimation for supercritical Galton-Watson processes. Statist. Probab. Lett. 50 331-342

22. Tamura, Roy N. and Boos, Dennis D..(1986). Minimum Hellinger distance estimation for multivariate location and covariate, Journal of the American Statistical Association, 81 223-229.

23. Tamura, Roy N., Faries, D.E., Anderson, J.s., and Heiligenstein, J.H. (1994). A case sttudy of an adaptive clinical trialin the treatment of out-patients with depressive disorder. Journal of the American Statistical Association, 89 768-776.

24. Vidyashankar (1994). Large deviation rates for branching processes in fixed and random environments, Thesis, Iowa State University.

25. 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

26. Wei, L.J. (1979). The generalized Pólya's urn design for sequential medical trials. The Annals of statistics, 7 291-296.

27. Wei, L.J. (1988). Exact two-sample permutation tests based on the randomized play-the-winner rule, Biometrika , 75, no. 3 603-606.

28. Zelen, M. (1969). Play the Winner Rule and the Controlled Clinical Trial. American Statistical Association Journal, 64 pp. 131-146.