BATCH SEQUENTIAL DESIGN FOR A NONLINEAR ESTIMATION PROBLEM

Werner G. MÜLLER
Benedikt M. PÖTSCHER

Forschungsbericht/
Research Memorandum No. 259

November 1989

Abstract

A method for constructing batch sequential designs for a nonlinear problem is presented. The limiting behaviour is investigated for a special model and simulation results are provided. As a by-product an optimal sampling theorem for martingale difference sequences is proved.

Eine Methode zur Konstruktion 'batch'-sequentieller Versuchspläne für ein nichtlineares Problem wird vorgestellt. Das asymptotische Verhalten wird für ein bestimmtes Modell untersucht und Simulationsergebnisse werden geliefert. Als Nebenprodukt wird ein Theorem der optimalen Auswahl von Martingaldifferenzenfolgen bewiesen.

Batch Sequential Design for a Nonlinear Estimation Problem

WERNER G. MÜLLER, BENEDIKT M. PÖTSCHER

Institute for Advanced Studies Vienna [1] , University of Technology Vienna [2]

I. Introduction

As is well known (see e.g. Fedorov, 1972) the optimal design measure in a nonlinear setup depends upon the unknown parameter, which is to be estimated.

Two solutions for this circular problem are suggested in the literature:

[a] the construction of the optimal design based on a prior guess θ_0 for the value of the parameter vector θ, or

[b] sequential methods, where the parameter estimators are updated during the experiment and the design is then based on the current estimators.

Procedures of the latter type can be performed either as a batch sequential or as a fully sequential method (i.e. with batches of size one), where after each batch of observations the estimation and the design are updated. Although the fully sequential method is expected to be theoretically superior than a batch sequential procedure (with batchsize > 1), in a number of applications practical reasons can lead to a preference for the latter, as for instance taking observations in batches may be cheaper than taking them individually.

[1] IHS,Stumpergasse 56, 1060 Vienna, Austria

[2] Department of Econometrics, Operations Research and System Theory, Argentinierstraße 8, 1040 Vienna, Austria

II. A Nonlinear Design Problem

The model of interest is the regression equation:

$$y_i = \theta_1 x_i + \theta_2 x_i^2 + \epsilon_i \tag{1}$$

where y_i is the response variable, x_i is the design variable , θ_1 and $\theta_2 \neq 0$ are parameters and the error ϵ_i is independently identically distributed as $N(0,1)$. The design variable x_i is assumed to take its values in the interval $[-1,1]$ only.

The goal is to estimate the value x, which optimizes the response, i.e. to estimate the nonlinear function $g(\theta) = \frac{-\theta_1}{2\theta_2}$, leading to a nonlinear design problem.

This example has been frequently used in the literature (e.g. Ford & Silvey,1980, Ford,Titterington & Wu,1985, Wu,1985 and Chaloner,1988) to illustrate the performance of optimal design methods in a nonlinear setup, since it is simple enough to allow an analytical treatment. Batch sequential design in a different framework was investigated by Abdelbasit & Plackett, 1983.

A reasonable and widely used design criterion is the asymptotic variance of the ML-estimator $\hat{g}_N$ of $g(\theta)$. Since this asymptotic variance is given by

$$var(\hat{g}_N) = \frac{1}{4\theta_2^2} c_g^T J_N^{-1} c_g \tag{2}$$

the design criterion becomes [3]

$$\Phi(J_N, \theta) = c_g^T J_N^{-1} c_g \tag{3}$$

where $c_g = (-2\theta_2)\dfrac{\partial g(\theta)}{\partial \theta} = \begin{pmatrix} 1 \\ 2g(\theta) \end{pmatrix}$, and $J_N = \begin{pmatrix} \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 \\ \sum_{i=1}^N x_i^3 & \sum_{i=1}^N x_i^4 \end{pmatrix}$ is the information matrix for N observations. It is well known, that the non-sequential optimal design for this situation takes values only at $x=-1$ and $x=1$ with the following probabilities:

$$p(1) = \begin{cases} 1/2 + g(\theta) & \text{if } |g(\theta)| \leq 1/2 \\ 1/2 + g(\theta)^{-1}/4 & \text{if } |g(\theta)| > 1/2 \end{cases} \tag{4a}$$

[3] Ford et al.,1985 criticize this choice, since it relies on properties of non-sequential design, Wu,1985, however, gives some asymptotic justification.

and

$$p(-1) = \begin{cases} 1/2 - g(\theta) & \text{if } |g(\theta)| \le 1/2 \\ 1/2 - g(\theta)^{-1}/4 & \text{if } |g(\theta)| > 1/2 \end{cases} \tag{4b}$$

The fully sequential design is found by minimizing $\Phi(J_i, \hat{\theta}_{i-1})$ with respect to the design variable x_i, see Silvey,1980. The batch sequential design is obtained as follows: for the k-th batch the spectrum $(x_{k1}, \ldots, x_{kn_k})$ is found by minimizing the optimality criterion $\Phi(J_{N_k}, \hat{\theta}_{k-1})$, where $N_k = \sum_{j=1}^{k} n_j$, n_j is the batch length of the j-th batch, $n_j \ge 1$, and $\hat{\theta}_{k-1}$ is the ML-estimator for θ after the $k-1$-th batch.[4] Of course, if $n_j = 1$ for all j, this reduces to the fully sequential design. Since optimization of $\Phi(J_{N_k}, \hat{\theta}_{k-1})$ is computationally burdensome and requires algorithms from approximative design theory we consider in the following an alternative method, which can be viewed as an approximation to the batch sequential design.[5] We call this method quasi-batch sequential design, which is given by the rule:

$$\text{Choose } x_{ki} \text{ which minimizes } \Phi(J_{N_{k-1}+i-1}, \hat{\theta}_{k-1}) \tag{5}$$

Since miminization of Φ may be computationally awkward, we shall instead, following the literature, choose the next design point by maximizing the Gateaux-derivative of $\log \Phi$ in the direction $(x_{ki}, x_{ki}^2)'(x_{ki}, x_{ki}^2)$, rather than by (5).

III. The Quasi Batch Sequential Design

It is easy to see, that maximizing the Gateaux-derivative mentioned above with respect to $x_{ki} \in [-1, 1]$ is equivalent to maximizing the scalar function

$$d_{(N_{k-1}+i-1)}(x_{ki}) = \left((x_{ki}, x_{ki}^2) J_{N_{k-1}+i-1}^{-1} c_{\hat{g}(k-1)}\right)^2, \tag{6}$$

[4] For ease of notation the j-th element of the k-th batch is here denoted by x_{kj}, i.e. $x_{kj} = x_{N_{k-1}+j}$ in the original notation. For convenience, we keep both notations, as no confusion seems possible. We also use the convention $N_0 = 0$.

[5] Another possibility would be to search for the non-sequential optimal design for N_k observations, then to find the closest (e.g. w.r.t. least squares criterion) design, which still contains the design from the $k-1$ step.

where $c_{\hat{g}(k-1)} = (1, 2g(\hat{\theta}_{(k-1)}))'$. It has been shown by Ford & Silvey, 1980, that this function can have its maximum only at $x = -1$ or $x = 1$, which therefore are the only candidates for design points.

An explicit solution for the optimization problem (6) can now be given quite analogously as in Ford & Silvey,1980 for the fully sequential case. Among the first N_{k-1} observations of the experiment N_{k-1}^- will be taken at $x = -1$ and $N_{k-1}^+ = N_{k-1} - N_{k-1}^-$ at $x = 1$. Let $\bar{y}_{k-1}^-$ denote the mean of the N_{k-1}^- observations on y at $x = -1$; $\bar{y}_{k-1}^+$ is defined analogously. Then $\bar{y}_{k-1}^- = -\hat{\theta}_{1(k-1)} + \hat{\theta}_{2(k-1)}$, $\bar{y}_{k-1}^+ = \hat{\theta}_{1(k-1)} + \hat{\theta}_{2(k-1)}$, and hence $2\hat{g}_{(k-1)} = (\bar{y}_{k-1}^- - \bar{y}_{k-1}^+)/(\bar{y}_{k-1}^- + \bar{y}_{k-1}^+)$.

After $N_{k-1} + i - 1$ observations the information matrix is given by

$$J_{N_{k-1}+i-1} = \begin{pmatrix} N_{k-1}+i-1 & (N_{k-1}+i-1)^+ - (N_{k-1}+i-1)^- \\ (N_{k-1}+i-1)^+ - (N_{k-1}+i-1)^- & N_{k-1}+i-1 \end{pmatrix}.$$

Simple algebra then yields

$$d_{(N_{k-1}+i-1)}(-1) \propto ((N_{k-1}+i-1)^+)^2(\bar{y}_{k-1}^+)^2, \tag{6a}$$

$$d_{(N_{k-1}+i-1)}(+1) \propto ((N_{k-1}+i-1)^-)^2(\bar{y}_{k-1}^-)^2. \tag{6b}$$

Hence, similarly as in Ford & Silvey, 1980 for the fully sequential case, the design rule for the quasi batch sequential design can be described as follows: The i-th observation of the k-th batch, i.e. the $(N_{k-1}+i)$-th observation, is taken at $+1$ or -1 according to whether

$$(N_{k-1}+i-1)^+|\bar{y}_{k-1}^+| \quad < \text{ or } \geq \quad (N_{k-1}+i-1)^-|\bar{y}_{k-1}^-|. \tag{7}$$

Of course the design rule (7) needs to be initialized. It suffices to fix the first two observations at $+1$ and -1. In case this allocation is chosen randomly it must not depend upon (ϵ_i). (Alternatively we can fix the entire first batch if $n_i > 1$.)

Note that the design rule (7) allows one to calculate the design of the k-th batch $(x_{k1}, \ldots, x_{kn_k})$ using only the first N_{k-1} observations. It may then be advantageous to rearrange $(x_{k1}, \ldots, x_{kn_k})$ prior to taking observations, such that only one change from 1 to -1 occurs in the batch. This avoids frequent setup changes, which may be costly.

IV. Asymptotics of Quasi Batch Sequential Design

The limiting behaviour of quasi batch sequential design is studied in this section. A proof of convergence of fully sequential designs in the particular example given above is provided in Ford & Silvey,1980. The argument given in that paper, however, seems to be incomplete, as those authors assume the validity of a law of large numbers for a certain optionally sampled sequence of i.i.d. variables without providing a proof. This lacuna in the proof is closed in Lemmas 2 and 3 in the Appendix. Given this law of large numbers, Ford & Silvey's,1980 proof can be generalized to provide a convergence proof for the more general case of quasi batch sequential designs. As corollaries consistency and asymptotic normality of the parameter estimators follow from standard martingale results. The proofs of the Theorem and Corollary can be found in the Appendix.

Theorem: Under the maintained assumptions the quasi batch sequential design converges to the non-sequential optimal design based on the true parameter value (given by (4)), if the batch lengths are bounded. I.e., if $n_i \leq \nu < \infty$ for all $i \geq 1$ (in particular if $n_i = \nu$), then

$$N^-/N \to p(-1) = |\theta_1 + \theta_2|/(|\theta_1 + \theta_2| + |\theta_2 - \theta_1|). \tag{8a}$$

$$N^+/N \to p(+1) = |\theta_2 - \theta_1|/(|\theta_1 + \theta_2| + |\theta_2 - \theta_1|). \tag{8b}$$

almost surely as $N \to \infty$.

The Theorem allows one to derive the asymptotic properties of the estimators $\hat{\theta}$ and $\hat{g}$ based on the quasi batch sequential design. Here $\hat{\theta}$ and $\hat{g}$ denote the estimators for a sample of size N and $\bar{y}^+ = \frac{1}{N^+}\sum_{i\in I_N^+} y_i$, $\bar{y}^- = \frac{1}{N^-}\sum_{i\in I_N^-} y_i$. The symbol $\to^d$ denotes convergence in distribution.

Corollary: Under the maintained assumptions the estimators $\hat{\theta}$ and $\hat{g}$ are strongly consistent. Furthermore, $\sqrt{N}(\hat{g} - g) \to^d N(0, (2\theta_2)^{-2} c_g' M^- c_g)$, where $M = \begin{pmatrix} 1 & p(1) - p(-1) \\ p(1) - p(-1) & 1 \end{pmatrix} = \lim N^{-1} J_N$ and M^- is any g-inverse.

If the optimal nonsequential design is not degenerate then also $\sqrt{N}(\hat{\theta} - \theta) \rightarrow^d N(0, M^{-1})$.

Inspection of the proofs shows that the Theorem and the Corollary also hold if the errors are not normally distributed but have a distribution with zero mean and unit variance.

V. Simulation Results

The small sample behaviour of quasi batch sequential design was studied by means of a simulation experiment. In order to make the results comparable with those reported in Ford and Silvey, 1980 for fully sequential designs the same simulation setup was used. Model (1) with $\theta = (1,4)$ and $\theta = (1,1)$, respectively, was used to generate samples of size 100, where the x_i were chosen by various design rules described below. The experiment was repeated 1000 times. At sample sizes 25,50,75 and 100 the same characteristics of the inference process as in Ford & Silvey,1980, e.g. the mean square error of $\hat{g}$, were calculated. We found no significant difference between the performance of quasi batch sequential and fully sequential design for sample size 50 and larger. (In fact an experiment with purely random design was not significantly inferior.)

For this reason we report only the results of the simulation for samples of size 25. As the results for $\theta = (1,1)$ are quite similar to the corresponding results for $\theta = (1,4)$ we report only the latter. In this case the response function has its maximum at $x = 0.25$.

The following design methods were studied: fully sequential design, quasi batch sequential design (5 batches of lenght 5)[6] and several non-sequential designs, namely (a) non-sequential optimal design under perfect information, i.e. $N_{25}^{-} = 15.625$, (b) non-informative non-sequential design, i.e. $N_{25}^{-} = 12.5$ and (c) pure random design, i.e. $N_{25}^{-} = 25\xi_j$ (ξ is a uniformly distributed random variable in

[6] In the first batch x was randomly chosen as ± 1 with probability $1/2$.

[0,1] and j is the number of the simulation run).

Table 1 presents the performance characteristics of the design methods given above ordered by the empirical mean square error of the estimator $\hat{g}$. As expected the non-sequential optimal design (a) performs best (least empirical m.s.e.). The difference in performance between fully and quasi batch sequential design is not substantial, while non-informative (b) and random (c) design are clearly dominated. These results support the asymptotic theory and seem to justify the use of quasi batch sequential procedures in expensive experiments.

design method	mean($\hat{g}_{25}$)	$m.s.e.(\hat{g}_{25}).10^5$	mean(N_{25}^-)	$var(N_{25}^-)$
non-sequential (a)	−0.1248	5.876	15.645	0.229
fully sequential	−0.1257	6.117	15.524	0.470
batch sequential	−0.1247	6.136	15.470	0.531
non-sequential (b)	−0.1265	6.990	12.510	0.250
random (c)	−0.1263	14.335	12.180	48.579

Table 1: Performance characteristics for different design techniques.

Figures 1 and 2 show the empirical distributions of the estimator $\hat{g}$ for N=5, 10 and 15 observations based on fully sequential and quasi batch sequential designs respectively. It can be seen that even for small samples the empirical distributions are not dramatically different, although the distributions in Figure 1 are slightly more concentrated.

Table 2 provides information about performance characteristics of quasi batch sequential procedures for different batch lengths. For samples of size $N = 25$ and $k = 5$ batches, different choices for the vector $n = (n_1, \ldots, n_5)$ were investigated.

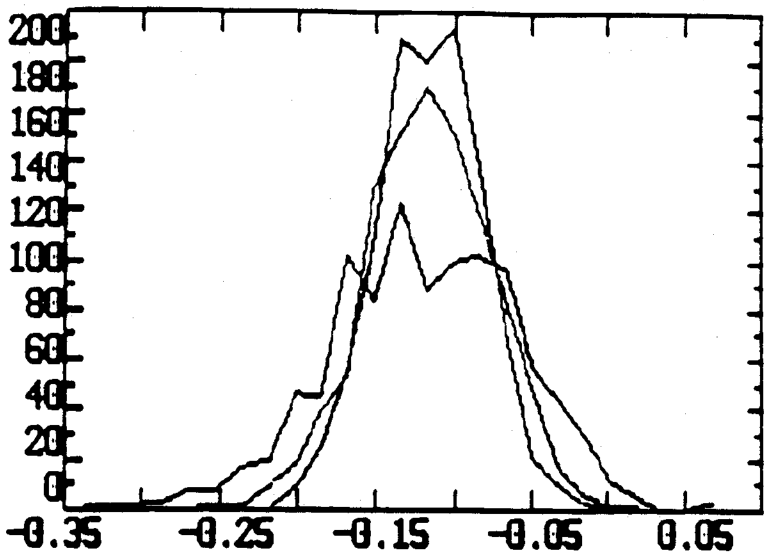

Figure 1: empirical distribution of $\hat{g}_5, \hat{g}_{10}, \hat{g}_{15}$ for fully sequential design.

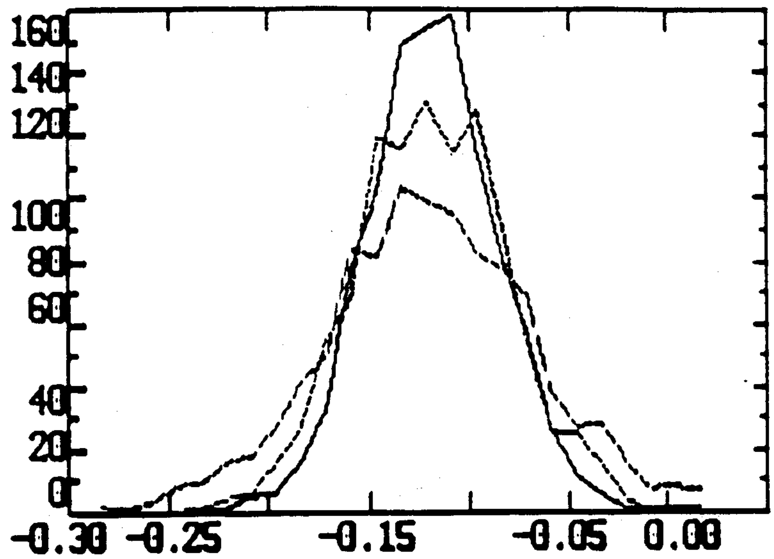

Figure 2: empirical distribution of $\hat{g}_5, \hat{g}_{10}, \hat{g}_{15}$ for quasi batch sequential design.

vector n	mean($\hat{g}_{25}$)	$m.s.e.(\hat{g}_{25}).10^5$	mean(N_{25}^-)	$var(N_{25}^-)$
$(5,5,5,5,5)$	-0.1257	6.136	15.52	0.548
$(2,2,17,2,2)$	-0.1245	6.229	15.40	0.612
$(5,5,5,5,5)^*$	-0.1247	6.234	15.47	0.531
$(17,2,2,2,2)^*$	-0.1253	6.267	15.51	0.484
$(2,2,2,2,17)$	-0.1238	6.391	15.48	1.246
$(2,17,2,2,2)$	-0.1257	6.422	15.49	0.620
$(17,2,2,2,2)$	-0.1236	6.483	15.41	0.371
$(2,2,2,17,2)$	-0.1244	6.638	15.42	0.944

Table 2: Performance characteristics for different quasi batch sequential designs.

The first batch was initialized as described above, except the cases marked by an asterisk, for which perfect information was used. The simulation showed no clear indication for the preference of any choice of vector n in this example. All the results for the empirical mean square error remained within the information gain of one additional information.

APPENDIX

Let $(\Omega, \mathcal{A}, P)$ denote the probability space on which all r.v. are defined.

Lemma 1: Under the assumptions of Section II, the quasi batch sequential design given by (7) satisfies: $N^+ \to \infty$ and $N^- \to \infty$ a.s. as $N \to \infty$.

Proof: It is sufficient to prove the Lemma for $N = N_k$. Suppose for some $\omega \in \Omega$ we have

$$N_k^-(\omega) = n \quad \text{for all } k \geq k^*(\omega). \tag{A.1}$$

Then condition (7) implies

$$|\sum_{i \in I_{N_k}^+} y_i(\omega)| < |\sum_{i \in I_{N_k}^-} y_i(\omega)| = const(\omega) < \infty \quad \text{for all } k \geq k^*(\omega).$$

It follows that

$$\limsup_{k\to\infty} |\sum_{i\in I^+_{N_k}} y_i(\omega)| = \limsup_{k\to\infty} |\sum_{i\in I^+_{N_{k^*(\omega)}}} y_i(\omega) + \sum_{i=N_{k^*(\omega)}+1}^{N_k} y_i(\omega)| \le const(\omega),$$

Hence we have

$$\limsup_{k\to\infty} |\sum_{i=N_{k^*(\omega)}+1}^{N_k} y_i(\omega)| < \infty \qquad (A.2)$$

on the event defined by (A.1). On this event the expression $|\sum_{i=N_{k^*(\omega)}+1}^{N_k} y_i|$ coincides with $|\sum_{i=N_{k^*(\omega)}+1}^{N_k} (\epsilon_i + \theta_1 + \theta_2)|$. Since ϵ_i is i.i.d. (with finite variance) it follows, that $\limsup_{s\to\infty} |\sum_{i=1}^{s} (\epsilon_i + \theta_1 + \theta_2)| = \infty$ a.s.. Hence in view of (A.2) the event defined by (A.1) has probability zero. Since the event $\{ N_k^- < \infty \}$ is a countable union of the above events we arrive at a contradiction. Thus $N_k^- \to \infty$ a.s.. The result for N_k^+ is proved analogously.

q.e.d.

Lemma 2: Suppose the assumptions of section II hold and let the design be generated by (7). Then

$$\lim_{N\to\infty} \frac{1}{N^+} \sum_{i\in I_N^+} y_i = \theta_1 + \theta_2 \quad \text{and} \quad \lim_{N\to\infty} \frac{1}{N^-} \sum_{i\in I_N^-} y_i = \theta_2 - \theta_1 \quad a.s..$$

Proof: Almost surely both values 1 and -1 occur infinitely often in the design (x_i) in view of Lemma 1. Ignoring a subset of probability zero in the following, the stopping times

$$\mu_j = \min\{i \ge 1 : \mathrm{card}(I_i^+) = j\},$$

where $\mathrm{card}(A)$ denotes the cardinality of a set A, are well-defined and finite. Furthermore the event $\{\mu_j = k\}$ depends only on $(x_1, \ldots, x_k)$ and hence is measurable with respect to $(\epsilon_1, \ldots, \epsilon_{k-1})$ (and the initializing design points) in view of the design rule (7). Define $z_j = y_{\mu_j}$ and $B_s = \frac{1}{s}\sum_{j=1}^{s} z_j$. Then the sequence $\frac{1}{N^+}\sum_{i\in I_N^+} y_i(\omega)$ is a subsequence of the sequence $(B_s(\omega))$ (possibly repeating values of $B_s(\omega)$). Hence it suffies to prove $B_s \to \theta_1 + \theta_2$ a.s. as $s \to \infty$. Observe that $x_{\mu_j} = 1$ by construction of μ_j. Therefore $z_j = \theta_1 + \theta_2 + \epsilon_{\mu_j}$ and it remains to show that $\frac{1}{s}\sum_{j=1}^{s} v_j \to 0$ a.s., where $v_j = \epsilon_{\mu_j}$. Since (ϵ_i) is i.i.d. with mean zero and finite second moment and since the initializing design points are independent of (ϵ_i), (ϵ_i) is a martingale difference sequence with respect to

$$\mathcal{G}_i = \sigma(\{\epsilon_i, \ldots \epsilon_1\};\ \text{initializing design points})$$

and satisfies the moment conditions in Lemma 3. As mentioned above $\{\mu_j = k\} \in \mathcal{G}_{k-1}$ is satisfied and hence (v_j) is a martingale difference sequence with uniformly bounded second moment by Lemma 3. A standard law of large numbers for martingale difference sequences, see e.g. Gänssler & Stute,1977,Korr.6.7.2., establishes $\frac{1}{s}\sum_{j=1}^{s} v_j \to 0$ a.s. and thus completes the proof.

q.e.d.

Note that Lemma 2 does not only hold for quasi batch sequential designs but for any design (x_i) satisfying (a) the design contains an infinite number each of 1 and -1 a.s. and (b) x_i is measurable w.r.t. $\mathcal{G}_{i-1}$.

The following lemma, which may be of independent interest, shows that an optionally sampled martingale difference sequence is again a martingale difference sequence if the transforming stopping times are predetermined in a sense made precise below.

Lemma 3: Let $(\zeta_i)_{i\geq 1}$ be a martingale difference sequence w.r.t. to a filtration $(\mathcal{G}_i)$ satistfying $\sup_i E(|\zeta_i|/\mathcal{G}_{i-1}) \leq c$ for some real constant c. Let $(\mu_j)_{j\geq 1}$ be an increasing sequence of stopping times which are a.s. finite and satisfy $\{\mu_j = k\} \in \mathcal{G}_{k-1}$. Put $v_j = \zeta_{\mu_j}$. Then (v_j) is a martingale difference sequence and $\sup_j E(|v_j|) \leq c$ holds. Furthermore, if additionally $\sup_i E(|\zeta_i|^\alpha/\mathcal{G}_{i-1}) \leq d$, d a real constant, $\alpha > 0$, then $\sup_j E(|v_j|^\alpha) \leq d$.

Proof: Define $\mathcal{F}_j$ as the σ-field generated by $v_j, \ldots, v_1$ and $\mu_{j+1}, \ldots, \mu_1$. Then v_j is clearly $\mathcal{F}_j$-measurable. To show $E(v_j/\mathcal{F}_{j-1}) = 0$ proceed as follows: consider sets $C \in \mathcal{F}_{j-1}$ of the form

$$C = \{(v_{j-1}, \ldots, v_1) \in B; \mu_1 = k_1, \ldots, \mu_j = k_j\} \qquad (A.3)$$

where B is a Borel set and $k_1 \leq \cdots \leq k_j$ are integers. Rewriting C as

$$C = \{(\zeta_{k_j-1}, \ldots, \zeta_{k_1}) \in B; \mu_1 = k_1, \ldots, \mu_j = k_j\}$$

we see that $C \in \mathcal{G}_{k_j-1}$. Assuming finiteness of $E(|v_j|)$ for the moment we obtain

$$\int_C E(v_j/\mathcal{F}_{j-1})dP = \int_C v_j dP = \int_C \zeta_{k_j} dP = E(1_C E(\zeta_{k_j}/\mathcal{G}_{k_j-1})) = 0 \qquad (A.4)$$

using $C \in \mathcal{F}_{j-1}, C \in \mathcal{G}_{k_j-1}$ and the martingale difference property of ζ_i. To establish finiteness of $E(|v_j|)$ we obtain quite similarly

$$\int_C E(|v_j|/\mathcal{F}_{j-1})dP = \int_C |v_j|dP = \int_C |\zeta_{k_j}|dP = E(1_C E(|\zeta_{k_j}|/\mathcal{G}_{k_j-1})) \leq c < \infty \qquad (A.5)$$

Now (A.5) also holds for countable disjoint unions of sets of the form as described in (A.3). Since $\Omega = \bigcup_{i=1}^{\infty} D_i, \quad D_i \subseteq D_{i+1}$ where each D_i is a countable disjoint union of sets of the form (A.3) (define D_i by fixing B and taking union over all possible $k_1 \leq \cdots \leq k_j$ and then let B increase with i), we have

$$E(|v_j|) = \int_\Omega E(|v_j|/\mathcal{F}_{j-1})dP = \lim_{i\to\infty} \int_{D_i} E(|v_j|/\mathcal{F}_{j-1})dP \leq c < \infty. \qquad (A.6)$$

Next observe that the family of sets of the form (A.3) generate $\mathcal{F}_{j-1}$ and that this family is stable w.r.t. finite intersections. The signed measure κ defined by $\kappa(A) = \int_A E(v_j/\mathcal{F}_{j-1})dP$, $A \in \mathcal{F}_{j-1}$, coincides with the zero measure on this generator in view of (A.4), i.e. the positive and the negative parts of κ coincide on this generator. The measure κ has finite total variation as a consequence of (A.6) and hence the positive and negative parts coincide on all of $\mathcal{F}_{j-1}$ by a standard uniqueness result for measures, see e.g. Gänssler & Stute,1977,Th.1.4.10.. But this means $\kappa(A) = 0$ for all $A \in \mathcal{F}_{j-1}$, i.e. $E(v_j/\mathcal{F}_{j-1}) = 0$. The proof of the final statement of the lemma is similar to the proof of (A.6).

q.e.d.

We note that in case (ζ_i) is i.i.d. with mean zero and finite variance one can show similarly to (A.4) and (A.5) that $E(v_j^2/\mathcal{F}_{j-1}) = E(\zeta_j^2)$ and that also $E(v_j^2 1_{(|v_j|>\delta\sqrt{N})}/\mathcal{F}_{j-1}) = E(\zeta_j^2 1_{(|\zeta_j|>\delta\sqrt{N})})$ hold (if $\mathcal{G}_i$ depends only on $\zeta_i, \ldots, \zeta_1$, and, possibly, information independent of the process (ζ_i)). Hence (v_j) satisfies the conditional Lindeberg condition and the norming condition of Brown's functional central limit theorem. (In fact, (v_j) is i.i.d. in this case as can be shown by a modification of the proof of Lemma 3.)

Proof of the Theorem: Since the batch length is assumed to be bounded it is sufficient to prove the result for $N = N_k$ with $k \geq 1$. For r a natural number let ${}_rN$ denote the index at which the r-th change from 1 to -1 occurs in the design, i.e. ${}_rN$ is the r-th smallest index such that $x_{{}_rN} = 1$ and $x_{{}_rN+1} = -1$ holds. Let ${}_rN^-$ (${}_rN^+$) denote the number of observations taken at $x = -1$ ($x = +1$) within the first ${}_rN$ observations. Since $N_k^+ \to \infty$ and $N_k^- \to \infty$ a.s. as $k \to \infty$ by Lemma 1, the random variable ${}_rN$ is a.s. finite and ${}_rN^+$, ${}_rN^-$ are a.s. well defined. Of course ${}_rN^+$ and ${}_rN^-$ are a.s. positive for $r > 1$, and ${}_rN^+ \to \infty$, ${}_rN^- \to \infty$ and ${}_rN \to \infty$ a.s. as $r \to \infty$. Let $k(r)$ denote the number of batches preceeding the r-th change from $+1$ to -1, i.e. $k(r)$ is determined such that $N_{k(r)} \leq {}_rN < N_{k(r)+1}$ holds. As the batch length is assumed to be bounded we clearly have $k(r) \to \infty$ as $r \to \infty$. Furthermore, note that $N_{k(r)}^+ > 0$ and $N_{k(r)}^- > 0$ a.s. hold for $r > \nu/2$. Applying now (7) to the $({}_rN+1)$-th observation we obtain

$$\frac{{}_rN^-}{N_{k(r)}^-} \Big| \sum_{i \in I_{N_{k(r)}}^-} y_i \Big| \leq \frac{{}_rN^+}{N_{k(r)}^+} \Big| \sum_{i \in I_{N_{k(r)}}^+} y_i \Big| \quad a.s., \tag{A.7}$$

where $I_{N_{k(r)}}^-$ is the index set corresponding to the observations in the first $k(r)$ batches taken at $x = -1$ and $I_{N_{k(r)}}^+$ denotes its complement relative to $\{i : 1 \leq i \leq N_{k(r)}\}$. Since Lemma 1 shows that $N_{k(r)}^+ \to \infty$ and $N_{k(r)}^- \to \infty$ a.s. as $r \to \infty$ we obtain from Lemma 2

$$\lim_{r \to \infty} \frac{1}{N_{k(r)}^-} \Big| \sum_{i \in I_{N_{k(r)}}^-} y_i \Big| = |\theta_1 - \theta_2| \quad a.s., \tag{A.8}$$

and

$$\lim_{r \to \infty} \frac{1}{N_{k(r)}^+} \Big| \sum_{i \in I_{N_{k(r)}}^+} y_i \Big| = |\theta_1 + \theta_2| \quad a.s.. \tag{A.9}$$

Dividing (A.7) by ${}_rN$ we obtain using (A.8) and (A.9)

$$|\theta_1 - \theta_2| \liminf_{r \to \infty} {}_rN^- / {}_rN \leq |\theta_1 + \theta_2| \liminf_{r \to \infty} {}_rN^+ / {}_rN \quad a.s.$$

Applying the design rule (7) to the $({}_rN)$-th observation we get

$$\frac{{}_rN^-}{N^-_{k(r)-1}}\Big|\sum_{i\in I^-_{N_{k(r)-1}}} y_i\Big| > \frac{{}_rN^+-1}{N^+_{k(r)-1}}\Big|\sum_{i\in I^+_{N_{k(r)-1}}} y_i\Big| \quad a.s., \tag{A.10}$$

on the event $\{{}_rN = N_{k(r)}\}$ and

$$\frac{{}_rN^-}{N^-_{k(r)}}\Big|\sum_{i\in I^-_{N_{k(r)}}} y_i\Big| > \frac{{}_rN^+-1}{N^+_{k(r)}}\Big|\sum_{i\in I^+_{N_{k(r)}}} y_i\Big| \quad a.s., \tag{A.11}$$

on the event $\{{}_rN > N_{k(r)}\}$. Again dividing by ${}_rN$ we obtain in any case

$$|\theta_1-\theta_2|\liminf_{r\to\infty} {}_rN^-/{}_rN \geq |\theta_1+\theta_2|\liminf_{r\to\infty} {}_rN^+/{}_rN \quad a.s.$$

This gives

$$|\theta_1-\theta_2|\liminf_{r\to\infty} {}_rN^-/{}_rN = |\theta_1+\theta_2|\liminf_{r\to\infty} {}_rN^+/{}_rN \quad a.s. \tag{A.12}$$

Similarly we get

$$|\theta_1-\theta_2|\limsup_{r\to\infty} {}_rN^-/{}_rN = |\theta_1+\theta_2|\limsup_{r\to\infty} {}_rN^+/{}_rN \quad a.s. \tag{A.13}$$

Observing that ${}_rN = {}_rN^- + {}_rN^+$, it follows from (A.12) and (A.13) for $\theta_1 \neq 0$ (note that $\theta_2 \neq 0$)

$$\lim_{r\to\infty} {}_rN^-/{}_rN = |\theta_1+\theta_2|/(|\theta_1+\theta_2|+|\theta_2-\theta_1|) \quad a.s.$$

If $\theta_1 = 0$, dividing (A.7) as well as (A.10),(A.11) by ${}_rN^-$ gives

$$1 \leq \liminf_{r\to\infty} {}_rN^+/{}_rN^- \leq \limsup_{r\to\infty} {}_rN^+/{}_rN^- \leq 1 \quad a.s.$$

observing that $\theta_2 \neq 0$. Hence

$$\lim_{r\to\infty} {}_rN^-/{}_rN = 1/2 = |\theta_1+\theta_2|/(|\theta_1+\theta_2|+|\theta_2-\theta_1|) \quad a.s.$$

If furthermore ${}^rN^-$ and ${}^rN^+$ denote the r-th change from -1 to 1 in the design, we obtain analogously as above

$$\lim_{r\to\infty} {}^rN^-/{}^rN = |\theta_1+\theta_2|/(|\theta_1+\theta_2|+|\theta_2-\theta_1|) \quad a.s.$$

Observing that N_k^-/N_k is monotonic between changes completes the proof.

q.e.d.

Proof of the Corollary: Since $\hat{\theta} = \frac{1}{2}(\bar{y}^+ - \bar{y}^-, \bar{y}^+ + \bar{y}^-)$ consistency follows from Lemma 2. Next observe that $\sqrt{N}(\hat{\theta} - \theta) = (N^{-1}J_N)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N (x_i, x_i^2)'\epsilon_i$ and that $N^{-1}J_N$ converges to M in view of the Theorem. From a standard martingale central limit theorem applied to $(x_i, x_i^2)'\epsilon_i$, cf. e.g. Gänssler & Stute, 1977,Th.9.3.2., and the Cramèr-Wold device we obtain $\frac{1}{\sqrt{N}}\sum_{i=1}^N (x_i, x_i^2)'\epsilon_i \to^d N(0, M)$. Observe that the conditional Lindeberg condition is satisfied for a linear combination $(\alpha x_i + \beta x_i^2)\epsilon_i$, since ϵ_i is i.i.d. with finite second moment and

$$E((\alpha x_i + \beta x_i^2)\epsilon_i^2 I_{(|\alpha x_i + \beta x_i^2||\epsilon_i| > \delta\sqrt{N})}/\mathcal{G}_{i-1}) \leq (|\alpha| + |\beta|)^2 E(\epsilon_i^2 I_{(|\epsilon_i| > \frac{\delta\sqrt{N}}{|\alpha|+|\beta|})}).$$

Consequently the last claim in the Corollary follows, since the optimal nonsequential design is nondegenerate iff M is nonsingular. As an immediate consequence we obtain the asymptotic distribution for $\sqrt{N}(\hat{g} - g)$ in the nondegenerate case. Finally, observe that M is singular iff $p(1) = 1$ or $p(-1) = 1$, i.e. $g = \frac{1}{2}$ or $g = -\frac{1}{2}$. As the proof for both cases is similar we give the proof only for the case $g = \frac{1}{2}$. Rewrite $\sqrt{N}(\hat{g} - g)$ as $-\sqrt{N}\bar{y}^+/(\bar{y}^+ + \bar{y}^-)$. By Lemma 2 the denominator converges to $2\theta_2$ and $N^+/N \to p(1) = 1$ by the Theorem. It hence suffies to establish $\bar{y}^+\sqrt{N^+} \to^d N(0,1)$ as $c_g' M^- c_g = 1$ holds. Now $\bar{y}^+\sqrt{N^+} = \frac{1}{\sqrt{N^+}}\sum_{i \in I_N^+} \epsilon_i$ is obtained by a random change of time from $\frac{1}{\sqrt{N}}\sum_{j=1}^N v_j$, v_j defined as in the proof of Lemma 2. In view of Lemma 3 and the remarks following this lemma, (v_j) is a martingale difference sequence satisfying the conditional Lindeberg and the norming condition of Brown's functional central limit theorem, cf. e.g. Gänssler & Stute,1977,Th.10.1.12. Since the random change of time satisfies $N^+/N \to 1$ a.s. in view of the Theorem, the asymptotic normality follows from Brown's functional central limit theorem together with Theorem 17.1 of Billingsley,1968.

q.e.d.