TRANSFORMATION OF OBSERVATIONS IN STOCHASTIC APPROXIMATION

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

SAMI NAGUIB ABDELHAMID

1971

This is to certify that the

thesis entitled

TRANSFORMATION OF OBSERVATIONS IN
STOCHASTIC APPROXIMATION

presented by

Sami Naguib Abdelhamid

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

Václav Fabian
Major professor

Date October 11, 1971

O-7639

ABSTRACT

TRANSFORMATION OF OBSERVATIONS IN STOCHASTIC APPROXIMATION

By

Sami Naguib Abdelhamid

Let X_1 be a random variable. We consider the following general stochastic approximation procedure:

$$(1) \qquad X_{n+1} = X_n - a_n c_n^{-1} h(Y_n) \;, \quad n = 1,2,\ldots$$

where Y_n are random variables, a_n and c_n are positive numbers, and h is a Borel measurable transformation. With the choice h = the identity, (1) includes both the Robbins-Monro (RM) procedure and the Kiefer-Wolfowitz (KW) procedure. Fabian (1960, 1964) considered (1) with h = sign; we shall call (1) with h = sign procedure (F).

We study the asymptotic properties, the a.s. convergence and the asymptotic normality, of this proposed generalized procedure under some mild requirements on h and on the random variables $V_n = (Y_n - M_n(\mathcal{X}_n)$ with $M_n(\mathcal{X}_n) = E_{\mathcal{X}_n}[Y_n]$ where $\mathcal{X}_n = [X_1, X_2, \ldots, X_n]$. We confine our analysis to the case where V_n are conditionally, given $\mathcal{X}_n$, distributed according to a distribution function G which is symmetric around 0 and admits a density g. It is shown that $n^{\beta}(X_n - \theta)$ is asymptotically distributed as a normal random variable ξ; θ is the parameter to be approximated. The effect of using h in (1) is pointed out for the RM and the KW situations.

We consider a transformation h optimal if it minimizes $E\,\xi^2$ and we show, under some regularity conditions, that h is optimal if and only if it is equal to $-C(g'/g)$ (a.e. with respect to G) for a $C > 0$. With such an optimal transformation h, the surprising result, despite the very simple recurrence relation in (1), is that our optimal procedure is not only optimal within the class of stochastic approximation procedures considered but also it is an asymptotically efficient estimator within the general class of regular unbiased estimators. We also show that the RM procedure as well as the KW procedure are optimal if and only if the error random variables are normally distributed. As for procedure (F) we show it is optimal if and only if the error random variables have a double exponential distribution. For distributions which do not satisfy the regularity conditions, we show how one can design transformations that yield improved procedures.

Our results make it possible to study the asymptotic relative efficiency (A.R.E.) for different choices of h, and in particular we show that the A.R.E. of procedure (F) relative to the optimal procedure is the same as the A.R.E. of the sequential sign test relative to the sequential probability ratio test (SPRT) (cf. Groeneveld (1971)).

TRANSFORMATION OF OBSERVATIONS IN STOCHASTIC APPROXIMATION

By

Sami Naguib Abdelhamid

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1971

TO MY WIFE AND SONS

Mona, Hishan and Tariq

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor Václav Fabian, who introduced me to the area of stochastic approximation and guided me through this research. His careful criticism and valuable comments helped me avoid many wrong turnings. His suggestions have been the source of many of the results presented here and improved virtually all results, either in substance or style, almost beyond recognition.

In addition, among many others who helped, I would like to thank Professor Dennis Gilliland for his encouragement and the time he gave for some discussions and reading the material of this research; Mrs. Noralee Barnes, who excellently typed both the rough and final versions of this thesis, and whose speed, accuracy, cheerfulness and patience leave me, at times, unnerved; my wife for her great understanding, continuous encouragement, heroic patience and endurance, and the ideal care of me and my two sons while I was too busy with my studies and research.

Finally, I would like to thank the people of my country, the United Arab Republic, and its Government for the financial support during the first five years of my graduate study. Also I would like to thank the Department of Statistics and Probability at Michigan State University for the generous support, financial and otherwise, that I have enjoyed over the period of my studies for the Ph.D. degree.

TABLE OF CONTENTS

Chapter		Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND SUMMARY

Consider the stochastic approximation procedures of the form

$$X_{n+1} = X_n - a_n c_n^{-1} Y_n, \quad n = 1,2,\dots , \tag{1.1}$$

where X_n, Y_n are random variables, and a_n, c_n are positive numbers. This includes both the Robbins-Monro (1951) procedure (RM), and the Kiefer-Wolfowitz (1952) procedure (KW).

We shall study the effect of transforming the random variables Y_n in (1.1) into $h(Y_n)$ by a Borel measurable transformation h. Fabian (1960 and 1964) proposed that Y_n in (1.1) be replaced by sign (Y_n). He studied the almost sure convergence of that modified procedure in both the RM and the KW situations (See §2.3, §2.4), and he also discussed analoguous modifications in the multidimensional case. He indicated that those modified procedures behave better in some practical applications.

Motivated by this idea can we transform Y_n by means of a Borel measurable transformation h and improve thereby the speed of convergence? The answer is yes. Therefore instead of (1.1), we shall consider the following general stochastic approximation procedure:

$$X_{n+1} = X_n - a_n c_n^{-1} h(Y_n), \quad n = 1,2,\dots , \tag{1.2}$$

where h is a Borel measurable transformation.

We shall establish the asymptotic properties of this proposed generalized procedure, and then we characterize the optimal transformation h. With the choice h = sign we will refer to (1.2) as procedure (F).

Let us denote

$$\mathcal{X}_n = [X_1, X_2, \ldots, X_n],\ M_n(\mathcal{X}_n) = E_{\mathcal{X}_n}[Y_n],\ V_n = Y_n - M_n(\mathcal{X}_n).$$

We shall confine our analysis to the case where the random variables V_n are conditionally (given $\mathcal{X}_n$) distributed according to a symmetric distribution function G admitting a density function g. This requirement of symmetry is natural in the KW situation (See §2.7 below).

To be more specific let f be a Borel measurable function defined on the real line. The exact analytic form of f may be unknown, but it is assumed that f belongs to a rather general family of functions. The only available information about f is that at any level x, we can observe f(x) subject to a random error; that is, we can obtain an unbiased observation of f(x). In the RM situation, we try to approximate the unknown root of the equation $f(\theta) = 0$; and in the KW situation we want to approximate θ, the unknown point of minimum (or maximum) of a function f.

In the RM situation (cf., e.g., Chung (1954), Hodges and Lehmann (1956), Burkholder (1956)) it is known, under some regularity conditions, that $n^{\frac{1}{2}}(X_n - \theta)$ is asymptotically distributed as a normal random variable ξ with

(1.3) $$E\,\xi = 0, \text{ and } \operatorname{Var}\xi = \frac{a^2\sigma^2}{2a\,f'(\theta) - 1},$$

where σ^2 = Variance of G. Also in the KW situation (cf. Fabian (1968b) and also the references there) it is known, under some regularity conditions, that $n^{1/3}(X_n - \theta)$ is asymptotically distributed as a normal random variable ξ with

(1.4) $$E\,\xi = -\frac{(2/3)a\;c^2 f'''(\theta)}{4\;a\;f''(\theta) - 2/3}, \text{ and } \operatorname{Var}\xi = \frac{\sigma^2 a^2 c^{-2}}{4\;a\;f''(\theta) - 2/3}.$$

The effect of transforming Y_n into $h(Y_n)$ is that instead of using estimators Y_n of $f(X_n)$, we are using estimators $h(Y_n)$ of another function $\tilde{f}(X_n)$. If then $\tilde{f}$ has the same root as f and if the conditions guaranteeing the asymptotic normality are preserved the effect of h is to change $f'(\theta)$ into $\tilde{f}'(\theta)$ in (1.3) and σ^2 into another variance $\tilde{\sigma}^2$. Similarly in the KW situation instead of using estimators Y_n of $[f(X_n + c_n) - f(X_n - c_n)]$, we are using estimators $h(Y_n)$ of $\tilde{f}(X_n, c_n)$, say. If then $c_n^{-1}\tilde{f}(X_n, c_n)$ has the same behavior as $c_n^{-1}(f(X_n + c_n) - f(X_n - c_n))$ and if conditions guaranteeing the asymptotic normality are also preserved the effect of h is to change $(f''(\theta), f'''(\theta))$ into $(\tilde{f}''(\theta), \tilde{f}'''(\theta))$ in (1.4), and σ^2 into another variance $\tilde{\sigma}^2$.

In the RM situation $E\,\xi^2$ will be minimized if $\dfrac{\tilde{\sigma}^2}{(\tilde{f}'(\theta))^2}$ is minimized within the class of all h's for which the asymptotic normality is preserved. Similarly in the KW situation $E\,\xi^2$ will be minimized if $\dfrac{\tilde{\sigma}^2}{(\tilde{f}''(\theta))^2}$ is minimized within the class of all such h's which preserve the asymptotic normality. The derivatives of $\tilde{f}$ at θ can be easily determined and we obtain

$\tilde{f}^{(i)}(\theta) = f^{(i)}(\theta)H(h)$ with

$$H(h) = [\frac{d}{dt}\int h(t+v)G(dv)]_{t=0} \tag{1.5}$$

for $i = 1$ in the RM- and $i = 2,3$ in the KW situation.

We shall state conditions on h under which both the almost sure convergence to θ and the asymptotic normality are preserved. Within the class $\mathcal{C}$ of such h's we consider the optimal transformation which minimizes the second moment of the asymptotic distribution of $n^{\beta}(X_n - \theta)$. This will be shown to be equivalent to finding $h \in \mathcal{C}$ which maximizes $H(h)$. This leads, under some regularity conditions, to $h = -(g'/g)$ (or any positive constant multiple of $-(g'/g)$).

The surprising fact is that with such an optimal h, the stochastic approximation procedure is not only optimal within the class of stoachstic approximation procedures considered but also, in the RM case, X_n is an asymptotically efficient estimator of θ within the class of all regular unbiased estimators.

Knowing the optimal transformation, we show that the RM procedure as well as the KW procedure are optimal if and only if the error random variables are normally distributed. As for procedure (F), we show it is optimal if and only if the error random variables have a double exponential distribution.

One of the regularity conditions is $0 < I(g) = \int (g'(v)/g(v))^2 G(dv) < \infty$; if it is not satisfied, we show for some particular distribution (e.g., uniform and triangular) how one can design transformations which yield improved procedures.

$\tilde{f}^{(i)}(\theta) = f^{(i)}(\theta)H(h)$ with

$$H(h) = [\frac{d}{dt}\int h(t+v)G(dv)]_{t=0} \quad (1.5)$$

for $i = 1$ in the RM- and $i = 2,3$ in the KW situation.

We shall state conditions on h under which both the almost sure convergence to θ and the asymptotic normality are preserved. Within the class $\mathcal{C}$ of such h's we consider the optimal transformation which minimizes the second moment of the asymptotic distribution of $n^{\beta}(X_n - \theta)$. This will be shown to be equivalent to finding $h \in \mathcal{C}$ which maximizes $H(h)$. This leads, under some regularity conditions, to $h = -(g'/g)$ (or any positive constant multiple of $-(g'/g)$).

The surprising fact is that with such an optimal h, the stochastic approximation procedure is not only optimal within the class of stoachstic approximation procedures considered but also, in the RM case, X_n is an asymptotically efficient estimator of θ within the class of all regular unbiased estimators.

Knowing the optimal transformation, we show that the RM procedure as well as the KW procedure are optimal if and only if the error random variables are normally distributed. As for procedure (F), we show it is optimal if and only if the error random variables have a double exponential distribution.

One of the regularity conditions is $0 < I(g) = \int (g'(v)/g(v))^2 G(dv) < \infty$; if it is not satisfied, we show for some particular distribution (e.g., uniform and triangular) how one can design transformations which yield improved procedures.

Our results also make it possible to compare different transformations, and we do so for the "sign" transformation (which yields procedure (F)), and the identity transformation (which yields either the RM procedure or the KW procedure). We then study the asymptotic relative efficiency of procedure (F) relative to the optimal procedure and show it has the same asymptotic erlative efficiency of the sequential sign test relative to the sequential probability ratio test (SPRT), (cf. Groeneveld (1971)).

CHAPTER 2

ALMOST SURE CONVERGENCE OF THE MODIFIED STOCHASTIC APPROXIMATION PROCEDURES

2.1 Basic Assumptions and Notations:

All random variables are supposed to be defined on a probability space $(\Omega,\mathcal{F},P)$. Relations between random variables, including convergence, are meant with probability one. E denotes the expectation and E_T the conditional expectation given a random vector T. The real line is denoted by R and the indicator function of a set Λ by χ_Λ. If ξ_n is a sequence of random variables (or in particular numbers) we use the notation $O(\xi_n)$, and $o(\xi_n)$ for denoting sequences of random variables such that there is a constant K and a number sequence $r_n \to 0$ such that $|O(\xi_n)| \leq K|\xi_n|$, and $|o(\xi_n)| \leq r_n|\xi_n|$.

2.2 Remark:

The original results by Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) were generalized and strengthened by Blum (1954). He proved $X_n \to \theta$ in both the RM and KW situation, as described below.

2.3 Robbins-Monro (RM) Situation:

Here we assume that (1.1) holds with $c_n = 1$, $M_n(X_n) = f(X_n)$, where f is a Borel measurable function on R such that

(2.3.1) $$\sup_{-k<x-\theta<-\frac{1}{k}} f(x) < 0, \inf_{\frac{1}{k}<x-\theta<k} f(x) > 0 ,$$

for an (unknown) number θ, and every natural number k. Furthermore, there exist constants A, B such that

(2.3.2) $$|f(x)| \le A|x - \theta| + B, \text{ for all } x \in R.$$

$(a_n)_{n=1}^{\infty}$ is a positive sequence of numbers satisfying

(2.3.3) $$\sum_{n=1}^{\infty} a_n = \infty , \sum_{n=1}^{\infty} a_n^2 < \infty .$$

(2.3.4) $$E_{\chi_n}[V_n^2] \le \sigma^2$$

for a number σ and every natural number n.

2.4 <u>Kiefer-Wolfowitz (KW) Situation</u>:

We assume that f is a Borel measurable function on R satisfying

(2.4.1) $$\sup_{-k<x-\theta<-\frac{1}{k}} \overline{D} f(x) < 0 , \inf_{\frac{1}{k}<x-\theta<k} \underline{D} f(x) > 0,$$

for an unknown number θ, and every natural number k; where $\underline{D} f(x)$, and $\overline{D} f(x)$ denote the lower and upper derivative, respectively, of f at x. Furthermore, there exist constants A, B such that

(2.4.2) $$|f(x+1) - f(x)| < A|x - \theta| + B \text{ for all } x \in R .$$

The relation (1.1) holds with positive a_n, c_n satisfying

(2.4.3) $$c_n \to 0, \sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} a_n^2 c_n^{-2} < \infty ;$$

and the random variables Y_n satisfy

(2.4.4) $$M_n(\mathfrak{X}_n) = f(X_n + c_n) - f(X_n - c_n),$$

and

(2.4.5) $$E_{\mathfrak{X}_n}[V_n^2] \le \sigma^2$$

for a number σ and every natural number n.

Burkholder ((1956); Theorem 1) proved a convergence theorem which contains the convergence results for both the RM and the KW situations as special cases. The following is a rewriting of that theorem in terms of the procedure (1.2).

2.5 Almost Sure (a.s.) Convergence Theorem:

Let $\theta \in R$ and σ^2 be a positive number. Let (1.2) hold and $\tilde{M}_n$, S_n^2 be Borel measurable functions; $\tilde{M}_n(X_n) = E_{\mathfrak{X}_n}[h(Y_n)]$, $S_n^2(X_n) = E_{\mathfrak{X}_n}[h(Y_n) - \tilde{M}_n(X_n)]^2$. Suppose that

(2.5.1) if $\epsilon > 0$, $|x - \theta| > \epsilon$ and $n > n_0(\epsilon)$ then

$$(x - \theta)\tilde{M}_n(x) > 0 \ ;$$

(2.5.2) if $0 < \delta_1 < \delta_2 < \infty$ then $\sum_{n=1}^{\infty} a_n c_n^{-1}[\inf_{\delta_1 \le |x-\theta| \le \delta_2} [|\tilde{M}_n(x)|]] = \infty$;

(2.5.3) $$\sup_{n,x} \left[\frac{|\tilde{M}_n(x)|}{1 + |x|}\right] < \infty \ ;$$

(2.5.4) $$\sup_{n,x} S_n^2(x) \le \sigma^2 \ ;$$

(2.5.5) $$\Sigma\, a_n^2 c_n^{-2} < \infty \ .$$

Then

$$X_n \to \theta \ . \tag{2.5.6}$$

2.6 Remark:

Given h, the a.s. convergence of the procedure (1.2) holds if the conditions of the preceding theorem are satisfied irrespective of whether they are also satisfied for h equal to the identity. But since we are interested in the optimal choice of h it seems useful to investigate conditions on h, which guarantee that conditions of Theorem 2.5 are satisfied with this h if they are satisfied for h = the identity.

Henceforth the following two assumptions will be assumed to hold.

2.7 Assumption:

In the general procedure (1.2), let

$$Y_n = M_n(X_n) + V_n$$

where V_n are random variables conditionally (given X_n) distributed according to a distribution function G which is symmetric around 0 and admits a density g. The functions M_n are Borel measurable.

Here, we may remark that the requirement of symmetry is natural in the KW situation where Y_n is an unbiased estimator of $[f(X_n + c_n) - f(X_n - c_n)]$. The requirement of symmetry is then satisfied if the errors in estimating $f(X_n + c_n)$ and $f(X_n - c_n)$ are independent and identically distributed.

2.8 Assumption:

We assume that h is an odd Borel measurable transformation defined on R and nonnegative on $[0,\infty)$. In addition we assume that $\Psi(t) = \int h(t+v)g(v)dv = \int h(v)g(v-t)dv$ exists for all $t \in R$.

2.9 Lemma:

Assume that

(2.9.1) (i) $\liminf_{t\downarrow 0} t^{-1}\Psi(t) > 0$,

and either

(ii) h is nondecreasing;

or

(iii) g is continuous and nonincreasing on $[0,\infty)$, and h is bounded and continuous; furthermore $h(v) > 0$ for all $v > 0$.

Then (2.5.1) and (2.5.2) hold for h if (2.5.1), (2.5.2), and (2.5.3) hold for the identity transformation.

Proof:

From (i) we obtain, for some positive numbers Δ and ρ_0, that

$$(2.9.2) \qquad t^{-1}\Psi(t) \geq \rho_0 \quad \text{for all} \quad 0 < t < \Delta .$$

If h is nondecreasing then so is Ψ and thus (2.9.2) implies that $\inf\{\Psi(t), t > t_0\} > 0$ for every $t_0 > 0$. Therefore (i) and (ii) imply that if $0 < T_0 < T_1 < \infty$, then

$$(2.9.3) \qquad \inf\{\Psi(t);\ t \in [T_0,T_1]\} > 0 .$$

Now suppose (iii) holds; we shall prove that (2.9.3) holds in this case, too. Since h is odd and g is symmetric, $\Psi(t)$ can be written in the form:

$$(2.9.4) \qquad \Psi(t) = \int_0^\infty h(v)[g(v-t) - g(v+t)]dv, \ t \in R \ .$$

The integrand is nonnegative for $t \geq 0$, since $h(v) \geq 0$ for $v \geq 0$, by Assumption 2.8, and $g(v-t) - g(v+t) \geq 0$ for $v \geq 0$ and every $t \geq 0$. The latter is obvious from (iii) if $0 \leq t \leq v$; if $0 \leq v < t$ then $g(v-t) - g(v+t) = g(t-v) - g(v+t)$ which is again nonnegative by (iii). In particular $\Psi(t) \geq 0$ for $t \geq 0$. But suppose $\Psi(t) = 0$ for some $t > 0$. Then, since $h(v) > 0$ for all $v > 0$, $g(v-t) - g(v+t) = 0$ for almost all (Lebesgue) $v \geq 0$. The function $F(v) = g(v-t) - g(v+t)$, $v \geq 0$, is then continuous and thus identically zero. Moreover g is periodic with period $2t$; but since g is nonincreasing for $v > 0$, then $g \equiv$ constant on R. This is a contradiction to the fact that g is a density function. Hence $\Psi(t) > 0$ for all $t > 0$. Furthermore since h is bounded and continuous, then by the dominated convergence theorem Ψ is continuous on $[T_0,T_1]$, $0 < T_0 < T_1 < \infty$. But $[T_0,T_1]$ is compact, then Ψ achieves its minimum on $[T_0,T_1]$ and thus (2.9.3) holds.

Since Ψ is odd, (2.9.3) implies $\Psi(t)$ sign $(t) \geq 0$. (2.5.1), (2.5.2) and (2.5.3) hold for M_n since (2.5.1) - (2.5.3) hold when $h =$ the identity. $\tilde{M}_n(X_n) = \Psi(M_n(X_n))$ and thus (2.5.1) for M_n implies (2.5.1) for $\tilde{M}_n$. Further (2.9.3) and (2.9.2) imply that for every $T > 0$ there is an $\eta > 0$ such that

(2.9.5) $\quad |\Psi(t)| \geq \eta|t| \quad$ for all $\quad |t| < T$.

Suppose $0 < \delta_1 < \delta_2$. Then using (2.5.3) for M_n we obtain $|M_n(x)| < T$ for some $T > 0$ and all $|x| \leq \delta_1$. Thus $|\tilde{M}_n(x)| = |\Psi(M_n(x))| \geq \eta|M_n(x)|$ for some $\eta > 0$, and (2.5.2) for $\tilde{M}_n$ follows from (2.5.2) for M_n. Q.E.D.

2.10 Lemma:

Assume there exist constants K_1, K_2 such that

(2.10.1) $\quad |\Psi(t)| \leq K_1|t| + K_2 \quad$ for all $\quad t \in R$.

Then (2.5.3) holds for $\tilde{M}_n$ if it does for M_n.

Proof:

Since $\tilde{M}_n(x) = \Psi(M_n(x))$,

$$\left[\frac{|\tilde{M}_n(x)|}{1 + |x|}\right] \leq \frac{K_1|M_n(x)| + K_2}{1 + |x|} \quad ,$$

thus (2.5.3) is satisfied since M_n satisfies (2.5.3). Q.E.D.

2.11 Remark:

Concerning (2.5.4) we notice that it is satisfeid if h is a bounded transformation. We also add this remark: if h is bounded by a straight line, M_n is bounded, and G has a bounded second moment then it is easy to verify that (2.5.4) holds; furthermore (2.5.3) also holds provided that M_n satisfies (2.5.3).

In the following we shall show, under some additional regularity conditions, that (2.5.1) - (2.5.4) of Theorem 2.5 hold for some particular choices of h in (1.2).

2.12 **Example**:

Let $h(v) = \text{sign}(v)$, $v \in R$. If g is continuous at 0, $g(0) \neq 0$, and (2.5.1) - (2.5.3) hold for M_n, then (2.5.1) - (2.5.4) hold for $\tilde{M}_n$.

Proof:

First of all the sign function clearly satisfies Assumption 2.8. Conditions (2.5.3) and (2.5.4) follow from the boundedness of h. Since h is nondecreasing, Lemma 2.9 will imply (2.5.1) and (2.5.2) if $\lim_{t\downarrow 0} t^{-1}\Psi(t) > 0$. But from the continuity of g at 0 we obtain

$$t^{-1}\Psi(t) = 2t^{-1}\int_{-t}^{0} g(v)\,dv \to 2g(0) > 0;\ g(0) \neq 0. \qquad \text{Q.E.D.}$$

2.13 **Example**:

Let h be a truncation function; that is for a positive number T_0 let $h(v) = v$ if $|v| < T_0$ and $|h(v)| = T_0$ if $|v| \geq T_0$. Assume that (2.5.1) - (2.5.3) hold for M_n and let $\int_0^{T_0} g(v)\,dv > 0$. Then (2.5.1) - (2.5.4) hold for $\tilde{M}_n$.

Proof:

Again it is obvious that h satisfies Assumption 2.8, and since h is nondecreasing; Lemma 2.9 will imply (2.5.1) and (2.5.2) if we show $\liminf_{t\downarrow 0} t^{-1}\Psi(t) > 0$. But

$$\Psi(t) = \int_{-T_0}^{T_0} (t+v)g(v)\,dv + T_0[1 - G(T_0-t) - G(-T_0-t)] ;$$

that is

$$\Psi(t) = t[G(T_0) - G(-T_0)] + T_0[G(T_0+t) - G(T_0-t)].$$

Thus for $t > 0$, we obtain

$$t^{-1}\Psi(t) \ge G(T_0) - G(-T_0) = 2G(T_0) - 1 .$$

Since $2G(T_0) - 1 > 0$, then

$$\liminf_{t \downarrow 0} t^{-1}\Psi(t) \ge 2G(T_0) - 1 > 0.$$

Concerning (2.5.3) and (2.5.4), they follow from the boundedness of h. Q.E.D.

2.14 __Example__:

Let h and h' be bounded and $\int h'(v)g(v)dv > 0$. In addition let g be continuous and nonincreasing on $[0,\infty)$ and $h(v) > 0$ for all $v > 0$. If M_n satisfies (2.5.1) - (2.5.3) then (2.5.1) - (2.5.4) hold for $\tilde{M}_n$.

__Proof__:

Since h is bounded, (2.5.3) and (2.5.4) hold. Also since (2.5.1) - (2.5.3) hold for M_n and (2.9.1)-(iii) is satisfied, Lemma 2.9 will imply (2.5.1) and (2.5.2) if we show that $\lim_{t \downarrow 0} t^{-1}\Psi(t) > 0$. Because of the boundedness of h', it follows that

$$\Psi'(0) = \int h'(v)g(v)dv > 0$$

which implies the required result. Q.E.D.

CHAPTER 3

ASYMPTOTIC NORMALITY OF THE MODIFIED PROCEDURES

In this chapter we shall use the following 1-dimensional version of a theorem in Fabian (1968b).

3.1 Theorem:

Let $\mathcal{F}_n$ be a non-decreasing sequence of σ-fields, $\mathcal{F}_n \subseteq \mathcal{F}$. Suppose U_n, V_n, T_n, Γ_n, Φ_n are random variables, σ_0, Γ, $\Phi \in R$, and $\Gamma > 0$. Suppose Γ_n, Φ_{n-1}, V_{n-1} are $\mathcal{F}_n$-measurable, C, α, $\beta \in R$, and

(3.1.1) $$\Gamma_n \to \Gamma,\ \Phi_n \to \Phi,\ T_n \to T, \text{ or } E[|T_n - T|] \to 0 .$$

(3.1.2) $$E_{\mathcal{F}_n}[V_n] = 0,\ C > |E_{\mathcal{F}_n}[V_n^2] - \sigma_0^2| \to 0 ,$$

and with

(3.1.3) $$\sigma_{j,r}^2 = E[\chi_{[|V_j|^2 \geq rj^{\alpha}]} |V_j|^2] ,$$

let either

(3.1.4) $$\lim_{j\to\infty} \sigma_{j,r}^2 = 0 \quad \text{for every} \quad r > 0 , \qquad \text{or}$$

(3.1.5) $$\alpha = 1,\ \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \sigma_{j,r}^2 = 0 \quad \text{for every} \quad r > 0 .$$

Suppose $\beta_+ = \beta$ if $\alpha = 1$, and $\beta_+ = 0$ if $\alpha \neq 1$.

(3.1.6) $$0 < \alpha \leq 1,\ 0 \leq \beta,\ \beta_+ < 2\Gamma , \quad \text{and}$$

(3.1.7) $$U_{n+1} = (1 - n^{-\alpha}\Gamma_n)U_n + n^{-(\alpha+\beta)/2}\Phi_n V_n + n^{-\alpha-\beta/2}T_n \ .$$

Then the asymptotic distribution of $n^{\beta/2}U_n$ is normal with

(3.1.8) mean $= 2T(2\Gamma - \beta_+)^{-1}$, and variance $= \sigma_0^2\Phi^2(2\Gamma - \beta_+)^{-1}$.

(For the proof see Fabian (1968b).)

3.2 <u>Asymptotic Normality Theorem</u>:

Let α_0, a, c and β be positive numbers, γ be a non-negative number and $\zeta_0 \in R$. Consider the modified procedure (1.2) with

$$a_n = a/n,\ c_n = c/n^{\gamma},\ 0 \le \gamma < \tfrac{1}{2},\ n = 1,2,\ldots\ ,$$

and suppose that $X_n \to \theta$*).

(3.2.1) Let h be continuous a.e. with respect to G, Ψ' exist at 0 and $\Psi'(0) = H(h) > 0$;

(3.2.2) $\beta = 1 - 2\gamma$ and $a > \dfrac{\beta}{2\alpha_0 H(h)}$.

Set

(3.2.3) $$S^2(t) = \int[h(t+v) - \Psi(t)]^2 g(v)\,dv,\ S_0^2(h) = S^2(0);$$

and assume that

(3.2.4) the function S^2 is bounded by a number σ^2 and is continuous at 0.

In addition, let α_n, ζ_n be $\mathfrak{X}_n$-measurable random variables, and with $s = \beta/(2\gamma)$ if $\gamma \neq 0$; $s = 0$ otherwise, assume that

---

*) This will be satisfied if $\tilde{M}_n$ satisfies (2.5.1)-(2.5.3), since (2.5.4) will be implied by (3.2.4) (cf. Theorem 2.5).

(3.2.5) $$c_n^{-1} M_n(X_n) = \alpha_n (X_n - \theta) + \zeta_n c_n^s ;$$

(3.2.6) $$\alpha_n \to \alpha_0 \quad \text{and} \quad \zeta_n \to \zeta_0 .$$

Then the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ is normal with

(3.2.7) $$\text{mean} = -2a\ c^s H(h) \zeta_0 [2a\ \alpha_0 H(h) - \beta]^{-1}$$

and

(3.2.8) $$\text{variance} = a^2 c^{-2} s_0^2(h) [2a\ \alpha_0 H(h) - \beta]^{-1} .$$

Proof:

The proof will be established by verifying the conditions of Theorem 3.1. We have

(3.2.9) $$(X_{n+1} - \theta) = (X_n - \theta) - a_n c_n^{-1} \tilde{M}_n(X_n) + a_n c_n^{-1} Z_n ;$$

where

(3.2.10) $$Z_n = -(h(Y_n) - \tilde{M}_n(X_n)) .$$

Define the following $\mathfrak{X}_n$-measurable random variables

(3.2.11) $$H_n = \begin{cases} H(h) & \text{if } M_n(X_n) = 0 , \\ M_n^{-1}(X_n) \tilde{M}_n(X_n) & \text{if } M_n(X_n) \neq 0; \ n = 1,2,\ldots . \end{cases}$$

Then the term $a_n c_n^{-1} \tilde{M}_n(X_n)$ in (3.29) can be written as $a_n H_n c_n^{-1} M_n(X_n)$; and using (3.2.5) we obtain

$$a_n c_n^{-1} \tilde{M}_n(X_n) = a_n \alpha_n H_n (X_n - \theta) + a_n c_n^s H_n \zeta_n .$$

Since $a_n c_n^s = a\ c^s n^{-1-\beta/2}$ and $a_n c_n^{-1} = a\ c^{-1} n^{-\frac{1}{2}-\beta/2}$, we can rewrite (3.2.9) as

$$(3.2.12)\quad (X_{n+1} - \theta) = (X_n - \theta)(1 - a\,\alpha_n H_n n^{-1}) + a\,c^{-1} n^{-(1+\beta)/2} Z_n - a\,c^s n^{-1-\beta/2} H_n \zeta_n \;.$$

Apply Theorem 3.1 with

$\alpha = 1$, $\mathcal{F}_n = \mathcal{X}_n$, $\Gamma_n = a\,\alpha_n H_n$, $U_n = (X_n - \theta)$, $\Phi_n = a/c$, $V_n = Z_n$, and

$T_n = -a\,c^s H_n \zeta_n$.

Now $c_n^{-1} M_n(X_n) \to 0$ by (3.2.5) since $X_n \to \theta$, and this with (3.2.1) implies that $H_n \to H(h)$. Thus

$$\Gamma_n \to a\,\alpha_0 H(h),\; T_n \to -a\,c^s H(h)\zeta_0,\; \Phi_n \to \Phi = a/c \;.$$

Also from (3.2.3) and the continuity of S^2 at 0 we obtain

$$E_{\mathcal{X}_n}[Z_n] = S^2(M_n(X_n)) \to S_0^2(h) \;.$$

So we have shown that (3.1.1) and (3.1.2) of Theorem 3.1 are satisfied with $\Gamma = a\,\alpha_0 H(h)$, $\Phi = a/c$, $T = -a\,c^s H(h)\zeta_0$, $C = \sigma^2$ and $\sigma_0^2 = S^2(0) = S_0^2(h)$.

Concerning (3.1.4) we have

$$(3.2.13)\quad \sigma_{j,r}^2 = E[\chi_{[Z_j^2 \ge rj]} Z_j^2] = E[E_{\mathcal{X}_j}[\chi_{[Z_j^2 \ge rj]} Z_j^2]] \;;$$

the conditional expectations form a uniformly integrable sequence since they are dominated by $E_{\mathcal{X}_j} Z_j^2 = S^2(M_j(X_j)) \le \sigma^2$, by (3.2.4). Thus (3.1.4) will be verified if we show that the conditional expectations converge to 0. But the j-th conditional expectation is equal to $Q_j(M_j(X_j))$ where

(3.2.14) $$Q_j(t) = \int \xi(t,v)\chi_{[\xi(t,v)\geq rj]}G(dv) ,$$

with

$$\xi(t,v) = [h(t+v) - \Psi(t)]^2 \quad \text{for all} \quad t,v \in R.$$

The integrands in (3.2.14) are uniformly integrable as $t \to 0$; since they are dominated by $\xi(t,v)$ for all $v \in R$ and $\int\xi(t,v)G(dv) = S^2(t) \to S_0^2(h)$ as $t \to 0$. Thus to show $Q_j(t) \to 0$ as $t \to 0$ and $j \to \infty$, it is enough to show for almost all (with respect to G) $v \in R$, $\xi(t,v)\chi_{[\xi(t,v)\geq rj]} \to 0$ as $t \to 0$ and $j \to \infty$; and for this it is enough to show for almost all (with respect to G) $v \in R$, $\xi(t,v) \to h^2(v)$ as $t \to 0$. Let Λ be the set of points at which h is discontinuous. Λ has a probability zero under G. By the continuity of h on $R - \Lambda$, and since by (3.2.1) Ψ is continuous at 0, then $\xi(t,v) \to h^2(v)$ as $t \to 0$ for all $v \in R - \Lambda$. This completes the proof of (3.1.4). The measurability condition follows from the definition of Γ_n, Φ_{n-1} and V_{n-1}. Also, by (3.2.2), $2a\,\alpha_0 H(h) > \beta$. Hence the conclusion of the theorem follows by Theorem 3.1. Q.E.D.

3.3 Remark:

Because of our interest in the question of optimality of the transformation h in the modified procedure (1.2), the preceding theorem is stated in terms of the behavior of M_n, (see conditions (3.2.5) and (3.2.6)), with sufficient conditions on h in order to guarantee the asymptotic normality of (1.2). But the theorem can also be applied to the modified procedure directly because this procedure can be written as the original procedure with a change of the meaning of Y_n's and with h equal to the identity transformation.

3.4 Remark:

A sufficient condition for (3.2.4) is that $(h(t+v) - \Psi(t))^2$ are uniformly integrable for t in a neighborhood of 0.

3.5 Theorem: (KW situation)

Let f, Y_n satisfy the conditions of the KW situation in §2.4 with possibly the exception of (2.4.5). Let f'' exist and be continuous in a neighborhood of θ with $f''(\theta) = M > 0$. Consider the modified procedure with a transformation h which is continuous a.e. with respect to G; further assume that (2.5.1) - (2.5.3) are satisfied for $\tilde{M}_n$ if they are for M_n. Let Ψ' exist at 0 with $\Psi'(0) = H(h) > 0$ and

(3.5.1) $$a_n = \frac{a}{n}, \quad c_n = \frac{c}{n^\gamma}, \quad \gamma = \frac{1}{4}, \quad a > [8M\, H(h)]^{-1} .$$

In addition let $S^2(t) = \int [h(t+v) - \Psi(t)]^2 g(v)\,dv$, $t \in R$, be bounded by a constant σ^2 and continuous at 0. Then $X_n \to \theta$ and the asymptotic distribution of $n^{1/4}(X_n - \theta)$ is normal with

(3.5.2) $$\text{mean} = 0, \text{ and variance} = a^2 c^{-2} S_0^2(h)[4M\, a\, H(h) - \tfrac{1}{2}]^{-1} .$$

Proof:

It is known (cf. Burkholder (1956)) and can be easily shown that $M_n(X_n) = f(X_n + c_n) - f(X_n - c_n)$ satisfy conditions (2.5.1) - (2.5.3) and thus by Theorem 2.5 we conclude that $X_n \to \theta$, since donditions (2.5.4) and (2.5.5) are satisfied, too. The rest of the conclusion of the theorem follows simply by verifying the conditions of Theorem 3.2. The only conditions of Theorem 3.2 which are not directly assumed in our theorem are (3.2.5), (3.2.6) and (3.2.2). Let $\Delta(x,c_n) = f(x+c_n) - f(x-c_n)$, $x \in R$. Let $I = [\theta - \epsilon, \theta + \epsilon]$ be an interval on which f'' exists and is continuous. Define

$$\varphi(x) = \begin{cases} (x-\theta)^{-1} f'(x) - f''(\theta) & \text{for } x \in I \text{ with } \varphi(\theta) = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5.3}$$

Then by expanding $\Delta(x,c_n)$ as a function of c_n and substituting for $f'(x)$ from (3.5.3), it follows that

$$c_n^{-1}\Delta(x,c_n) = 2(x-\theta)[f''(\theta) + \varphi(x)] + \eta(x,c_n)c_n \tag{3.5.4}$$

where $\varphi(x) \to 0$ as $x \to \theta$ and $\eta(x,c_n) \to 0$ if $c_n \to 0$ and $x \to \theta$. Obviously φ and $\eta(\cdot,c_n)$, the latter as defined by (3.5.4), are Borel measurable functions. Then

$$c_n^{-1}M_n(X_n) = \alpha_n(X_n - \theta) + \zeta_n c_n$$

with $\mathcal{I}_n$-measurable $\alpha_n = 2[f''(\theta) + \varphi(X_n)]$ and $\zeta_n = \eta(X_n,c_n)$; further $\alpha_n \to 2f''(\theta) = 2M$ and $\zeta_n \to 0$. Thus (3.2.5) and (3.2.6) hold with $\alpha_0 = 2M$, $\zeta_0 = 0$ and $s = 1$, since $\beta = \frac{1}{2}$. Condition (3.2.2) then follows from (3.5.1). Hence conditions of Theorem 3.2 are satisfied. This completes the proof. Q.E.D.

3.6 <u>Theorem</u>: (KW situation)

In Theorem 3.5 let (3.5.1) be replaced by

$$a_n = \frac{a}{n},\ c_n = \frac{c}{n^\gamma},\ \gamma = \frac{1}{6} \text{ and } a > [6M\, H(h)]^{-1}. \tag{3.6.1}$$

Moreover let f''' exist and be continuous in a neighborhood of θ. Then $X_n \to \theta$ and the asymptotic distribution of $n^{1/3}(X_n - \theta)$ is normal with

(3.6.2) $$\text{mean} = -(2/3)ac^2 H(h) f'''(\theta)[4M\ a\ H(h) - \tfrac{2}{3}]^{-1}, \text{ and}$$
$$\text{variance} = a^2 c^{-2} S_0^2(h)[4M\ a\ H(h) - \tfrac{2}{3}]^{-1} .$$

Proof:

It is again easy to conclude, by Theorem 2.5, that $X_n \to \theta$, and then it remains to verify (3.2.5), (3.2.6) and (3.2.2) of Theorem 3.2 in order to complete the proof of the theorem. Let $\Delta(x,c_n) = f(x+c_n) - f(x-c_n)$, $x \in R$. Let $I = [\theta - \epsilon, \theta + \epsilon]$ be an interval on which f''' exists and is continuous. Similarly as in the proof of the preceding theorem and with the same φ, we obtain that

(3.6.3) $$c_n^{-1} M_n(X_n) = \alpha_n(X_n - \theta) + \zeta_n c_n^2$$

with $\mathfrak{X}_n$-measurable $\alpha_n = 2[f''(\theta) + \varphi(X_n)]$ and $\zeta_n = \frac{1}{3}[f'''(\theta) + \eta(X_n c_n)]$; further $\alpha_n \to 2f''(\theta) = 2M$ and $\zeta_0 = \frac{1}{3} f'''(\theta)$. Thus (3.2.5) and (3.2.6) hold with $\alpha_0 = 2M$, $\zeta_0 = \frac{1}{3} f'''(\theta)$, $\beta = \frac{2}{3}$ and $s = 2$. Condition (3.2.2) follows from (3.6.1). Hence conditions of Theorem 3.2 are satisfied. This completes the proof. Q.E.D.

The following theorem is presented here to cover a case treated in Albert and Gardner (1967) and the results below will be used later in Chapter 5 of this paper.

3.7 Theorem:

Let f_n be a sequence of Borel measurable functions defined on R such that (2.5.1) - (2.5.3) are satisfied for $M_n = f_n$ with

$c_n = 1$ and $f_n(\theta) = 0$. Let $d > 0$ and

(3.7.1) $$D_n(x) = \begin{cases} (x-\theta)^{-1} f_n(x) & \text{if } x \neq \theta, \\ d & \text{if } x = \theta;\ n = 1,2,\ldots \end{cases}$$

be continuously convergent at θ to d. Consider the modified procedure (1.2) with a transformation h which is continuous a.e. with respect to G, and let h be such that (2.5.1) - (2.5.3) are satisfied for $\tilde{M}_n$ if they are for M_n. Also let Ψ' exist at 0, $\Psi'(0) = H(h) > 0$ and

(3.7.2) $$a_n = \frac{a}{n}, \ a > [2d\, H(h)]^{-1}.$$

In addition let $S^2(t) = \int [h(t+v) - \Psi(t)]^2 g(v)dv$, $t \in R$, be bounded by a constant σ^2 and continuous at 0. Then $X_n \to \theta$ and the asymptotic distribution of $n^{\frac{1}{2}}(X_n - \theta)$ is normal with

(3.7.3) mean = 0 and variance = $a^2 S_0^2(h)[2ad\, H(h) - 1]^{-1}$.

Proof:

Under the given conditions we obtain, by applying Theorem 2.5, that $X_n \to \theta$. To obtain the rest of the conclusion of the theorem, we apply Theorem 3.2 for which we need only to verify (3.2.5), (3.2.6) and (3.2.2). We have

$$M_n(X_n) = f_n(X_n) = D_n(X_n)\ (X_n - \theta)\ .$$

Since $\gamma = 0$, then $\beta = 1$ and $s = 0$. Thus (3.2.5) and (3.2.6) hold with $\mathfrak{X}_n$-measurable $\alpha_n = D_n(X_n)$ and $\zeta_n = \zeta_0 = 0$; further $\alpha_n \to d_0 = d$, since $(D_n)_{n=1}^{\infty}$ is continuously convergent at θ to d. Condition (3.2.2) follows from (3.7.2). Hence the conclusion follows by Theorem 3.2. Q.E.D.

3.8 Theorem: (RM situation)

Let f, Y_n satisfy the conditions of the RM situation in §2.3 with possibly the exception of (2.3.4); further let f' exist at θ with $f'(\theta) > 0$. Consider the modified procedure (1.2) with a transformation h which is continuous a.e. with respect to G and for which (2.5.1) - (2.5.3) are satisfied for $\tilde{M}_n$ if they are for M_n. Also let Ψ' exist at 0, $\Psi'(0) = H(h) > 0$ and

(3.8.1) $$c_n = 1,\ a_n = \frac{a}{n},\ a > [2f'(\theta)H(h)]^{-1} .$$

In addition let $S^2(t) = \int[h(t+v) - \Psi(t)]^2 g(v)dv$, $t \in R$, be bounded by a constant σ^2 and continuous at 0. Then $X_n \to \theta$ and the asymptotic distribution of $n^{\frac{1}{2}}(X_n - \theta)$ is normal with

(3.8.2) $$\text{mean} = 0 \quad \text{and} \quad \text{variance} = a^2 S_0^2(h)[2a\ f'(\theta)H(h) - 1]^{-1} .$$

Proof:

It is known (cf. Burkholder (1956)) that M_n satisfies (2.5.1) - (2.5.3). In Theorem 3.7 let $f_n = f$, then it follows that conditions of Theorem 3.7 are satisfied with $d = f'(\theta) > 0$. Hence the conclusion of the theorem follows by Theorem 3.7. Q.E.D.

CHAPTER 4

SPECIAL CASES AND RESULTS

4.1 Introduction:

In this chapter we show that the a.s. convergence (cf. Fabian (1960)) and the asymptotic normality of procedure (F) follow as a special case of the corresponding results for the modified procedure.

Then we give the optimal choice of a and c for the modified procedure in both the KW and the RM situation. Also in this chapter we point out the effect of taking more observations at each stage.

Throughout the rest of this paper, we write $N(\mu_1,\mu_2^2)$ to denote a normal random variable with mean $= \mu_1$ and variance $= \mu_2^2$, and we also write $T_n \xrightarrow{\mathcal{L}} \xi$ if T_n is asymptotically ξ-distributed.

4.2 The a.s. convergence and the asymptotic normality of procedure (F):

In §4.3, §4.4 and §4.5 we consider the modified procedure (1.2) with the choice h = the sign transformation, and we assume that g is continuous at 0 and $g(0) \neq 0$. Hence $H(h) = 2g(0) > 0$; furthermore, the sign function is continuous except at 0 and $S^2(t) = 1 - \Psi^2(t) < \infty$ for all $t \in R$, Since Ψ is continuous at 0, then so is S^2 with $S_0^2(h) = 1$. Moreover if (2.5.1) - (2.5.3) are satisfied for M_n, then (see Example 2.12) $\tilde{M}_n$ satisfy

(2.5.1) - (2.5.4). Thus the statements in Results 4.3, 4.4, and 4.5 follow from Theorems 3.5, 3.6, and 3.8, respectively.

4.3 Result: (KW situation)

Let f, Y_n satisfy the conditions of the KW situation in §2.4 with possibly the exception of (2.4.5). Let f''' exist and be continuous in a neighborhood of θ with $f''(\theta) = M > 0$. Let

(4.3.1) $$a_n = \frac{a}{n},\ c_n = \frac{c}{n^\gamma},\ \gamma = 1/4, \text{ and } a > [16M\ g(0)]^{-1}.$$

Then $X_n \to \theta$ and

(4.3.2) $$n^{\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{L}} N(0, a^2c^{-2}[8Ma\ g(0) - \tfrac{1}{2}]^{-1}).$$

4.4 Result: (KW situation)

In Result 4.3 let (4.3.1) be replaced by

(4.4.1) $$a_n = \frac{a}{n},\ c_n = \frac{c}{n^\gamma},\ \gamma = 1/6 \text{ and } a > [12M\ g(0)]^{-1}.$$

In addition let f'' exist and be continuous in a neighborhood of θ. Then $X_n \to \theta$ and

(4.4.2) $$n^{1/3}(X_n - \theta) \xrightarrow{\mathcal{L}} N\left(-\frac{(4/3)a\ c^2 g(0) f'''(\theta)}{8Ma\ g(0) - 2/3}, \frac{a^2c^{-2}}{8Ma\ g(0) - 2/3}\right).$$

4.5 Result: (RM situation)

Let f, Y_n satisfy the conditions of the RM situation in §2.3 with possibly the exception of (2.3.4). In addition let f' exist at θ, $f'(\theta) > 0$ and

(4.5.1) $$c_n = 1,\ a_n = \frac{a}{n},\ a > [4f'(\theta)g(0)]^{-1}.$$

Then $X_n \to \theta$ and

$$(4.5.2) \qquad n^{\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{L}} N(0,\ a^2[4f'(\theta)a\ g(0) - 1]^{-1}) .$$

4.6 Remark:

With the choice h = the identity transformation, the reader may easily check that the a.s. convergence and the asymptotic normality of the original RM and KW procedures follow as special cases of the corresponding results for the modified procedure provided Assumption 2.7 holds and $\int v^2 g(v)\,dv < \infty$.

4.7 The optimal choice of (a,c) in the KW situation:

Let ξ be a normal random variable with mean and variance given by (3.6.2). Then

$$(4.7.1) \qquad E\,\xi^2 = \frac{(4/9)a^2H^2(h)f'''^2(\theta)}{(4Ma\ H(h) - 2/3)^2}\ c^4 + \frac{a^2S_0^2(h)}{(4Ma\ H(h) - 2/3)}\ c^{-2} .$$

We shall find the optimal values of (a,c), i.e. values for which $E\,\xi^2$ is minimized with other quantities being fixed. We assume that $f'''(\theta) \neq 0$. To find the value $c(a)$ of $c > 0$ that minimizes the R.H.S. of (4.7.1) for a fixed $a > [6M\ H(h)]^{-1}$, let

$$A = \frac{a^2S_0^2(h)}{4Ma\ H(h) - 2/3}\ , \quad B = \frac{(4/9)a^2H^2(h)f'''^2(\theta)}{(4Ma\ H(h) - 2/3)^2} \quad \text{and}$$

$$\Psi(c) = A\ c^{-2} + B\ c^4 .$$

Differentiating Ψ with respect to c we get

$$\Psi'(c) = -2A\ c^{-3} + 4B\ c^3 .$$

Thus $[c(a)]^6 = \frac{A}{2B}$; that is

(4.7.2) $$[c(a)]^6 = \frac{9}{8} \frac{(4Ma\ H(h) - 2/3)}{H^2(h) f'''^2(\theta)} S_0^2(h) .$$

With $c = c(a)$, (4.7.1) becomes

(4.7.3) $$E\ \xi^2 = [3H^2(h) f'''^2(\theta)]^{1/3} \frac{a^2[S_0^2(h)]^{2/3}}{(4Ma\ H(h) - 2/3)^{4/3}} .$$

Now the R.H.S. is minimized by the choice

(4.7.5) $$a = \frac{1}{2M\ H(h)} ,$$

which can be easily verified. Therefore the optimal choice of (a,c) is given by

(4.7.6) $$(a,c) = \left(\frac{1}{2M\ H(h)} , [\tfrac{3}{2} f'''^2(\theta)[S_0^2(h)/H^2(h)]]^{1/6}\right).$$

With this optimal choice of (a,c), (4.7.1) becomes

(4.7.7) $$E\ \xi^2 = (3/16M^2)[(9/4) f'''^2(\theta)]^{1/3}[S_0^2(h)/H^2(h)]^{2/3} .$$

4.8 <u>The optimal choice of a in the RM situation</u>:

Let ξ be a normal random variable with mean and variance given by (3.8.2), then

(4.8.1) $$E\ \xi^2 = \frac{a^2 S_0^2(h)}{2a\ f'(\theta)H(h) - 1} .$$

The optimal choice of a, say a^*, is the value of a which minimizes $E\ \xi^2$ with other quantities fixed. Thus one can easily check that

(4.8.2) $$a^* = \frac{1}{f'(\theta)H(h)} .$$

With $a = a^*$, (4.8.1) becomes

$$(4.8.3) \qquad E\,\xi^2 = S_0^2(h)/(f'^2(\theta)H^2(h)) \ .$$

4.9 Remark:

We notice that the unpleasant feature of the optimal values of a and c is that they depend on values, $(f'(\theta), f''(\theta), f'''(\theta))$, which are, in general, unknown. But the value of a in the RM situation and the value of (a,c) in the KW situation can be estimated during the approximation process and fed back into the procedure.

For the original RM procedure, Venter (1967) used a procedure, (later generalized by Fabian (1968b)), which estimates the optimal value of a. Recently in Fabian (1971) a procedure was described which itself estimates the optimal value of a for a modified version of the original KW procedure. The same ideas can be used to obtain a procedure which estimates the optimal value of a or both (a,c) for the modified procedure.

4.10 Effect of taking m observations at each stage:

Suppose that an experimenter observes m random variables $Y_{n,1},\ldots,Y_{n,m}$ instead of one, Y_n, at stage n such that these m random variables are conditionally, given $\mathfrak{X}_n$, independently distributed according to G. Suppose he then uses $\frac{1}{m}\sum_{j=1}^{m} h(Y_{n,j})$ instead of $h(Y_n)$ in the modified procedure (1.2). The conditional expectation of the average will be the same as that of $h(Y_n)$ and the conditional variance will only be changed by a factor of $(1/m)$ and it is easy to see, under the conditions of Theorem 3.2, that this will result in changing the variance of the asymptotic distribution by the factor $(1/m)$. Thus, in the KW situation, under the conditions of Theorem 3.6 we obtain

$$(4.10.1)\qquad n^{1/3}(X_n - \theta) \overset{\mathcal{L}}{\to} N\left(-\frac{(2/3)a\,c^2 f'''(\theta)H(h)}{4Ma\,H(h) - 2/3}, \frac{a^2c^{-2}S_0^2(h)}{m(4Ma\,H(h) - 2/3)}\right).$$

On the other hand if the experimenter wishes to continue the modified procedure for nm stages rather than using averages, then by Theorem 3.6 we have

$$(4.10.2)\qquad (nm)^{1/3}(X_{nm} - \theta) \overset{\mathcal{L}}{\to} N\left(-\frac{(2/3)a\,c^2 f'''(\theta)H(h)}{4Ma\,H(h) - 2/3}, \frac{a^2c^{-2}S_0^2(h)}{4Ma\,H(h) - 2/3}\right).$$

Let $f'''(\theta) \neq 0$. Then in (4.10.2) the optimal choice of (a,c) (see (4.7.6)), is given by

$$(4.10.3)\qquad (a,c) = \left(\frac{1}{2M\,H(h)}, \left[\frac{3}{2} f'''^2(\theta)\left[\frac{S_0^2(h)}{H^2(h)}\right]\right]^{1/6}\right),$$

while the optimal choice of (a,c) in (4.10.1) is given by

$$(4.10.4)\qquad (a,c) = \left(\frac{1}{2M\,H(h)}, \left[\frac{3}{2} f'''^2(\theta)\left[\frac{S_0^2(h)}{m\,H^2(h)}\right]\right]^{1/6}\right).$$

Let ξ_1, ξ_2 be the normal random variables on the R.H.S. of (4.10.1) and (4.10.2), respectively. Then with the corresponding optimal choice of (a,c), one can easily check that

$$E\,\xi_1^2 = E\,\xi_2^2 .$$

Therefore using average of m independent observations at each stage is asymptotically equivalent to continuation of the modified procedure for nm stages. The only effect is in decreasing the optimal value of c by a factor of $(1/m)^{1/6}$. Also, in the RM situation, under the conditions of Theorem 3.8, and by using average we obtain

$$(4.10.5) \qquad (nm)^{\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{L}} N(0,\ a^2 S_0^2(h)/(2a\ f'(\theta)H(h) - 1)),$$

while if the modified procedure is continued for nm stages, then from Theorem 3.8 we obtain

$$(4.10.6) \qquad (nm)^{\frac{1}{2}}(X_{nm} - \theta) \xrightarrow{\mathcal{L}} N(0, a^2 S_0^2(h)/(2f'(\theta)a\ H(h) - 1)) .$$

Hence from (4.10.5) and (4.10.6) we see that, in the RM situation, asymptotically there is no effect.

CHAPTER 5

OPTIMAL TRANSFORMATIONS

5.1 Introduction:

We have seen, in Chapter 3, that the asymptotic results for the modified procedure (1.2) in both the RM situation and the KW situation are special cases of the situation described in Theorem 3.2. The second moment of the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ can be written (see (3.2.7) and (3.2.8)) as

$$(5.1.1) \qquad \frac{4a^2c^{2s}\zeta_0^2}{(2a\alpha_0 - (\beta/H(h)))^2} + \frac{a^2c^{-2}S_0^2(h)}{(2a\alpha_0 H(h) - \beta)} .$$

If $\zeta_0 \neq 0$ then, with the optimal choice of (a,c), (5.1.1) becomes

$$(5.1.2) \qquad \frac{\beta^2}{\alpha_0^2}((s+1)/s)^3[4s\ \zeta_0^2]^{1/s+1}[\frac{s}{\beta(s+2)}]^{s+2/s+1}[S_0^2(h)/H^2(h)]^{s/s+1}.$$

Both (5.1.1) and (5.1.2) cover the KW situation, but in the RM situation $c = 1$, $\zeta_0 = 0$, and $\beta = 1$, thus (5.1.1) reduces to

$$(5.1.3) \qquad \frac{a^2S_0^2(h)}{2a\alpha_0 H(h) - 1} ,$$

which with the optimal choice of a becomes

$$(5.1.4) \qquad [S_0^2(h)/\alpha_0^2H^2(h)] .$$

All the above expressions depend on h only through $H(h)$ and $S_0^2(h)$. We notice that $H(bh) = bH(h)$ and $S_0^2(bh) = b^2S_0^2(h)$ for any positive number b, and so a change from h to bh does not affect (5.1.2) and (5.1.4). But if we change a to (a/b), then the stochastic approximation procedure will not change and (5.1.1), (5.1.3) will be unchanged, too. This shows that it is enough to consider transformations with $S_0^2(h) = 1$. Then any one of the above expressions is minimized by the choice h which maximizes H(h).

Let $\mathcal{K}$ be the family of all Borel measurable transformations h such that h satisfies Assumption 2.8, $S_0^2(h) = 1$, and H(h) can be computed by differentiating under the integral sign; that is

$$0 < H(h) = \Psi'(0) = \int h(v)(-g'(v))dv .$$

5.2 <u>Theorem</u>*):

Let the density g have a derivative a.e. with respect to G. In addition let

(5.2.1) $$0 < I(g) = \int (g'(v)/g(v))^2 dG(v) < \infty ,$$

and set $\Gamma = [I(g)]^{\frac{1}{2}}$. Suppose that $h_0 = -\frac{1}{\Gamma}(g'/g)$ a.e. with respect to G and $h_0 \in \mathcal{K}$. Then within $\mathcal{K}$, H(h) is maximized by h^* if and only if $h^* = h_0$ a.e. with respect to G.

<u>Proof</u>:

We have

$$H(h) = \int h(v)(-g'(v))dv .$$

---

*) The author wishes to thank Professor Peter Bickel for calling his attention to the possibility of using the Schwarz inequality in proving the theorem after it was proved by a different method.

By Schwarz inequality, we obtain

$$(5.2.2) \qquad H^2(h) \le (\int h^2(v)g(v)dv)\cdot\int(g'(v)/g(v))^2 g(v)dv = \int(g'^2(v)/g(v))dv.$$

Since $0 < \int(g'(v)/g(v))^2 g(v)dv < \infty$ and $\int h^2(v)g(v)dv = 1$ then equality holds in (5.2.2) if and only if (cf. Theorem 3.5 of Rudin (1966, p. 61)) there exists a non zero constant K such that

$$h = K(-g'/g) \quad \text{a.e. with respect to } G.$$

To have $H(h) > 0$ we must have $K > 0$ and to satisfy $S_0^2(h) = 1$ we have to have $K = \frac{1}{\Gamma}$. This completes the proof. Q.E.D.

5.3 Remark:

A sufficient condition for $H(h)$ to be equal to $\int h(v)(-g'(v))dv$ is that for a certain neighborhood, N, of 0 the family $\{h(x)\frac{g(x-t)-g(x)}{t\, g(x)};\ t \in N-\{0\}\}$ be uniformly integrable with respect to G.

5.4 Definition:

Suppose that $0 < I(g) < \infty$, $h^* = -\frac{1}{\Gamma}(g'/g)$ a.e. with respect to G, and $h^* \in \mathcal{H}$. In addition suppose that the modified procedure is used with $h = h^*$ for which $X_n \to \theta$ and $n^{\beta/2}(X_n - \theta)$ has asymptotic distribution as given in Theorem 3.2. Then we shall call h^* the optimal transformation.

5.5 Remark:

An optimal transformation h has the following property. For any competitor $h \in \mathcal{H}$ to which Theorem 3.2 also applies,

the second moment of the asymptotic distribution is larger than that for h^* unless $h = h^*$ a.e. with respect to G.

A modified procedure using an optimal transformation will be called optimal procedure.

5.6 Asymptotic efficiency of optimal stochastic approximation procedures; the RM situation:

The surprising fact is that, in the RM situation, the optimal stochastic approximation procedures are not only optimal within the class of approximation procedures considered but also they are asymptotically efficient within the general class of regular unbiased estimators of θ, the parameter to be estimated. This is true despite the very simple recurrence relation that generates the approximation sequence X_n.

In more detail, we show that the variance of the asymptotic distribution of $n^{\frac{1}{2}}(X_n - \theta)$ corresponds to the Cramer-Rao lower bound for the variance of an unbiased estimator based on the first n observations. We shall consider the following two cases:
(i) a straight line case; $f(x) = d(x-\theta)$ with $d > 0$ known,
(ii) a case of sequence of functions $f_n(\theta)$ which are known except for θ. This second case was considered and studied in Albert and Gardner (1967).

5.7 The straight line case:

Let

(5.7.1) $$Y_n = f(X_n) + V_n, \quad n = 1,2,\ldots$$

where $f(x) = d(x-\theta)$ with θ unknown but assumed to belong to some interval, Θ; d positive and known. In addition assume that $V_n = Y_n - f(X_n)$ are independent and distributed according to G which satisfies the conditions stated in Theorem 5.2.

The information contained in $[Y_1,Y_2,\ldots,Y_n]$ is the same as that contained in $[Z_1,Z_2,\ldots,Z_n]$, where $Z_j = Y_j + dX_j$, $j = 1,2,\ldots,n$. Then the Cramer-Rao lower bound for the variance of any unbiased regular estimator of θ, based on the first n observations, is given (cf. Theorem 4.1.1 of Zacks (1971), p. 186) by $I_n^{-1} = n^{-1}(d\Gamma)^{-2}$. Thus the asymptotic efficiency of X_n is 1, since $n^{\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{L}} N(0,(d\Gamma)^{-2})$.

5.8 <u>A case of a sequence of functions</u>:

Let

$$Y_n = f_n(\theta) + V_n \ , \quad n = 1,2,\ldots \tag{5.8.1}$$

be observations on known functions f_n except for θ which is assumed to belong to some interval, Θ. Let the error random variables $V_n = Y_n - f_n(\theta)$ be independent and distributed according to G which satisfies the conditions of Theorem 5.2. Furthermore for each n let f_n have the same unique root $\theta \in \Theta$, f_n exist at θ and $f_n'(\theta) \to d$ where d is positive and known. Also let f_n satisfy the conditions stated in Theorem 3.7.

Albert and Gardner have shown (cf. Theorem 5.2 of Albert and Gardner (1967), p. 68) that in this case X_n will be asymptotically efficient among regular unbiased estimators of θ if the variance of the asymptotic distribution of $n^{\frac{1}{2}}(X_n - \theta)$ is

$(d\Gamma)^{-2}$ which is true for our X_n generated by the optimal procedure. Hence our optimal procedure is asymptotically efficient within the class of regular unbiased estimators of θ.

Albert and Gardner (1967; see Chapter 5 there) tried to increase the efficiency of the RM type procedure which they used in their monograph by making transformation of the parameter space Θ. Their procedure stayed less efficient except when the error random variables are normally distributed.

5.9 Theorem:

The KW procedure as well as the RM procedure are optimal if and only if the error random variables are normally distributed.

Proof:

The procedure is optimal if and only if $(-g'/g)(v) = Cv$ with a constant $C > 0$, and this is true if and only if G is a normal distribution. Q.E.D.

5.10 Theorem:

Procedure (F) is optimal if and only if the error random variables have a double exponential distribution.

Proof:

Procedure (F) is optimal if and only if $(-g'/g)(v) = C \text{ sign } (v)$ with a constant $C > 0$ and this is true if and only if G is a double exponential distribution. Q.E.D.

5.11 Some examples of new optimal procedures:

In the following we give four examples of new optimal procedures which are different from the original RM and KW procedures. The first three examples fall under the case of Example 2.14.

5.12 Example:

Let G have a Cauchy density function

$$g(v) = \frac{1}{\pi}\frac{1}{1+v^2} \quad , v \in R .$$

(Recall that random errors with this density function are not allowed by the RM procedure as well as the KW procedure, since both of them require the existence of the second moment of G.) One can check that $\Gamma^2 = 5/2$. Let

$$h_1(v) = -\frac{1}{\Gamma} g'(v)/g(v) = 2(2/5)^{\frac{1}{2}} \frac{v}{1+v^2} , v \in R.$$

This is an odd bounded monotone transformation with a bounded first derivative. Hence one can easily check that $h_1 \in \mathcal{K}$, and h_1 maximizes $H(h)$. Furthermore h_1 preserves the a.s. convergence (see Example 2.14) and it satisfies the conditions of Theorem 3.2. Thus h_1 is an optimal transformation.

5.13 Example:

Let G have a Student's t-density function given by

$$g(t) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} (1 + t^2/\nu)^{-(1+\nu)/2}, \ t \in R, \ \nu > 0,$$

where ν here stands for the degrees of freedom. With some manipulation one can check that

$$h_2(t) = -\frac{1}{\Gamma} (g/(t)/g(t)) = C_\nu \frac{t}{\nu+t^2} , \ t \in R,$$

where C_ν is a positive constant for which

$$\int_{-\infty}^{\infty} C_\nu^2 \left(\frac{t}{\nu+t^2}\right)^2 g(t)\,dt = 1 .$$

Again this transformation, h_2, has the same properties as h_1 in §5.12, above. Hence h_2 is an optimal transformation.

5.14 Example:

Let G have a logistic density function given by:

$$g(v) = \frac{1}{2(1+\cosh(v))}, \quad v \in R.$$

We leave it to the reader to check (see §2.14) that the optimal transformation is given by

$$h_3(v) = C_3 \frac{\sinh(v)}{1+\cosh(v)}, \quad v \in R ,$$

where $C_3 = \frac{\sqrt{3}}{2}$.

5.15 Example:

Let G have a density function given by:

$$g(v) = \begin{cases} \dfrac{C_0}{\sqrt{2\pi}} \; e^{-v^2/2} , & \text{if } |v| < T \\ \dfrac{C_0}{\sqrt{2\pi}} \; e^{-T|v|+(T^2/2)} , & \text{if } |v| \geq T , \end{cases}$$

where C_0 and T are positive constants.

This g behaves like a normal density for small v, and then like a double exponential for large v. It follows that

$$-\frac{1}{\Gamma}\,(g'(v)/g(v)) = \begin{cases} C_4 v & \text{if } |v| < T \\ C_4 T \operatorname{sign}(v) & \text{if } |v| \geq T, \end{cases}$$

where C_4 also depends on T. Denoting this transformation by h_4 we see that h_4 is an odd, bounded and nondecreasing transformation which satisfies Assumption 2.8. Thus $h_4 \in \mathcal{N}$ and $h_4 = -\frac{1}{\Gamma} g'/g$ a.e. maximizes $H(h)$. Also h_4 preserves the a.s. convergence (see Lemmas 2.9 and 2.10) and it satisfies the conditions of Theorem 3.2. Hence h_4 is an optimal transformation.

5.16 Modified procedures by means of suitable transformations:

In Theorem 5.2 we characterized the optimal transformation for a given density function g which has a derivative a.e. with respect to G and for which

$$0 < I(g) = \int (g'(v)/g(v))^2 dG(v) < \infty.$$

What about those distributions for which $I(g) = 0$ or ∞? For example, if g is a constant symmetric density function then $g' = 0$ a.e. and thus $I(g) = 0$. Also if g' itself is constant we have $I(g) = \infty$ (let g be triangular on $(-1,1)$).

In the following we give two cases to show how to design transformations that yield better behaved procedures; but a unique optimal transformation does not exist in these cases.

5.17 A case in which $g' = 0$ a.e. (G):

Let G be the uniform distribution on $(-b,b)$, $b > 0$; that is

$$g(v) = \begin{cases} 1/2b & , -b < v < b, \\ 0 & \text{otherwise.} \end{cases}$$

Let $r > 1$, and consider a transformation of the form:

$$h_r(v) = \begin{cases} \sqrt{2r+1}\ |v/b|^r \operatorname{sign}(v) & , |v| < b \\ \sqrt{2r+1}\ \operatorname{sign}(v) & , |v| \geq b \ . \end{cases}$$

This is a bounded nondecreasing transformation which satisfies Assumption 2.8 and one can check (see Lemmas 2.9 and 2.10) that this transformation preserves the a.s. convergence of the procedure (1.2) as well as its asymptotic normality. Let f, a_n be as in Theorem 3.8 with $a > [2f'(\theta)(2r+1)^{\frac{1}{2}}]^{-1}$. Then for the RM procedure, by using (4.8.3) with the choice $h =$ the identity we obtain

(5.17.1)
$$E\,\xi^2 = \frac{1}{f'^2(\theta)} \frac{b^2}{3} ,$$

while with the choice $h = h_r$, for $r > 1$, (4.8.3) gives

(5.17.2)
$$E\,\xi^2 = \frac{1}{f'^2(\theta)} \frac{b^2}{(2r+1)} .$$

Therefore we see that the ratio of (5.17.1) to (5.17.2) is always larger than 1 for $r > 1$. Taking, e.g. $r = 10$, will result in decreasing $E\,\xi^2$ by the factor $\frac{1}{7}$.

5.18 <u>A case in which</u> $g' =$ <u>constant</u> a.e. (G):

In particular we consider the density

$$g(v) = \begin{cases} 1 - |v| & , -1 < v < 1 \\ 0 & \text{otherwise} . \end{cases}$$

Let $0 < r < 1$, and consider the transformation

$$(5.18.1) \qquad h_r(v) = \begin{cases} C_r[\dfrac{1}{(1-|v|)^r} - 1]\text{sign}(v), & -1 < v < 1 , \\ C_r \text{ sign}(v) & \text{otherwise} , \end{cases}$$

where $C_r = \frac{1}{r}[(1-r)(2-r)]^{\frac{1}{2}}$. This transformation satisfies Assumption 2.8 and it can be checked by direct calculations that

$$\Psi'(0) = H(h) = 2 \frac{2-r}{1-r} .$$

Also we can verify that this transformation preserves the a.s. convergence and satisfies conditions of Theorem 3.2. Let f, a_n satisfy the conditions in Theorem 3.8 with $a > [2f'(\theta)H(h_r)]^{-1}$, then using (4.8.3) with the choice $h = h_r$, it follows that

$$(5.18.2) \qquad E\,\xi^2 = \frac{1}{f'^2(\theta)} \frac{(1-r)}{4(2-r)} ,$$

while with the choice h = the identity (4.8.3) gives

$$(5.18.3) \qquad E\,\xi^2 = \frac{1}{f'^2(\theta)} \frac{1}{6} .$$

To see the effect of this transformation in the KW situation, let f, a_n, c_n satisfy the conditions stated in Theorem 3.6 with $a > [6M\, H(h_r)]^{-1}$. Then with the optimal choice of (a,c) provided that $f'''(\theta) \neq 0$, it follows (see §4.7), with the choice $h = h_r$, that

$$(5.18.4) \qquad E\,\xi^2 = \frac{3}{16M^2} [(9/4)f'''^2(\theta)]^{1/3} [\frac{1-r}{4(2-r)}]^{2/3} ,$$

while for the KW procedure (h = the identity) we have

$$(5.18.5) \qquad E\,\xi^2 = \frac{3}{16M^2} [(9/4)f'''^2(\theta)]^{1/3} [\frac{1}{6}]^{2/3} .$$

Thus we are able to make the ratio of (5.18.5) to (5.18.4) (or (5.18.3) to (5.18.2)) as large as we wish by choosing r close to 1.

CHAPTER 6

ASYMPTOTIC RELATIVE EFFICIENCY OF THE MODIFIED PROCEDURES

In Chapter 3 we have shown, under some regularity conditions, that, for a $\beta > 0$, $n^{\beta/2}(X_n - \theta) \xrightarrow{\mathcal{L}} \xi$ where ξ is normally distributed with certain mean and variance. As a reasonable measure of comparison between different procedures we use $E\,\xi^2$ either with a and c chosen optimally or with c fixed and a chosen to minimize $E\,\xi^2$.

Our results in Chapters 3, 4 and 5 make it possible to compare different transformations, and we do so for the identity transformation and the sign transformation. We also compare the sign transformation to the optimal transformation.

To the knowledge of the present author there have not been any study of the asymptotic relative efficiency (A.R.E.) of the RM and KW procedures relative to procedure (F) except some simulation study by Springer (1969).

6.1 Definition of the asymptotic relative efficiency (A.R.E.):

Let $\mathcal{P}_{h_1}$ and $\mathcal{P}_{h_2}$ denote two modified procedures with transformations h_1 and h_2 such that $n^{\beta/2}(X_n - \theta) \xrightarrow{\mathcal{L}} \xi_{h_1}$ and $n^{\beta/2}(X_n - \theta) \xrightarrow{\mathcal{L}} \xi_{h_2}$, respectively. Then the A.R.E. of $\mathcal{P}_{h_1}$ relative to $\mathcal{P}_{h_2}$ is taken to be

$$e(\theta_{h_1};\theta_{h_2}) = \frac{E[\xi_{h_2}^2]}{E[\xi_{h_1}^2]} .$$

θ_{h_1} is called more efficient than θ_{h_2} if $e(\theta_{h_1};\theta_{h_2}) > 1$.

6.2 Comparison of some transformations:

In Table I, we list, for four different cases, the values of the second moment of the asymptotic distribution for the choices h = the identity, h = sign and h = the optimal transformation (see Definition 5.4). Also we list the A.R.E. of h = sign relative to h = the identity, and the A.R.E. of h = sign relative to the optimal transformation.

In cases (1) and (2) g is assumed to be continuous at 0 with $g(0) \neq 0$ and $\sigma^2 = \int v^2 g(v)\,dv < \infty$; further in cases (3) and (4) g is assumed to satisfy the conditions of Theorem 5.2 and to be continuous at 0 with $g(0) \neq 0$. Moreover in cases (1) and (3) let f, Y_n satisfy the conditions of the RM situation in §2.3 with $a_n = \frac{a}{n}$, and in addition let f' exist at θ and $f'(\theta) > 0$. Then with h = the identity (provided $a > [2f'(\theta)]^{-1}$) and also with h = sign *) (provided $a > [4f'(\theta)g(0)]^{-1}$), conditions of Theorem 3.8 are satisfied. In cases (2) and (4) let f, Y_n satisfy the conditions of the KW situation in §2.4 with $a_n = \frac{a}{n}$, $c_n = \frac{c}{n^\gamma}$, $\gamma = 1/6$ and in addition let f''' exist and be continuous in a neighborhood of θ, and $f''(\theta) = M > 0$. Then with h = the identity (provided $a > [6M]^{-1}$) and h = sign *)

---

*) See §4.2 and Results 4.4 and 4.5.

TABLE I

Case	Values of the second moment of the asymptotic distribution			A.R.E.
	Identity h_1	Sign h_2	Optimal *) $\tilde{h}_0 = -\frac{1}{\Gamma}(g'/g)$	
(1)	$\sigma^2/f'^2(\theta)$	$[4f'^2(\theta)g^2(0)]^{-1}$		$e(h_2;h_1) = 4g^2(0)\sigma^2$
**) (2)-(i)	$\sigma^2/6c^2M^2$	$[24c^2M^2g^2(0)]^{-1}$		$e(h_2;h_1) = 4g^2(0)\sigma^2$
(2)-(ii)	$\frac{3}{16M^2}[\frac{3}{2}f'''(\theta)\sigma^2]^{2/3}$	$\frac{3}{16M^2}[\frac{(3/2)f'''(\theta)}{4g^2(0)}]^{2/3}$		$e(h_2;h_1) = [4g^2(0)\sigma^2]^{2/3}$
(3)		$[4f'^2(\theta)g^2(0)]^{-1}$	$[f'^2(\theta)\Gamma^2]^{-1}$	$e(h_2;h_1) = \frac{4g^2(0)}{\Gamma^2}$
(4)		$\frac{3}{16M^2}[\frac{(3/2)f'''(\theta)}{4g^2(0)}]^{2/3}$	$\frac{3}{16M^2}[\frac{(3/2)f'''(\theta)}{\Gamma^2}]^{2/3}$	$e(h_2;h_1) = [\frac{4g^2(0)}{\Gamma^2}]^{2/3}$

*) See Definition 5.4.

**) In this case values are computed by direct manipulation of the corresponding results for the choices h_1 and h_2.

(provided $a > [12M\, g(0)]^{-1}$), conditions of Theorem 3.6 are satisfied. In case (2)-(i) $f'''(\theta) = 0$, and in case (2)-(ii) $f'''(\theta) \neq 0$. Optimal constants a, c are chosen except in case (2)-(i) where c is fixed. Thus the values of the second moment of the asymptotic distribution are obtained from (4.8.3) in cases (1) and (3) and from (4.7.7) in cases (2)-(ii) and (4).

6.3 Remark:

From case (3) in Table I, the A.R.E. of procedure (F) relative to the optimal procedure is given by $e(h_2;h_0) = 4g^2(0)/\int (g'^2(v)/g(v))dv$; furthermore if c is fixed and $h_0 = -\frac{1}{\Gamma}(g'/g)$ is an optimal transformation, then from case 2 in Table I we also obtain that $e(h_2;h_0) = 4g^2(0)/\int (g'^2(v)/g(v))dv$. Recently Groeneveld (1971) has shown that the A.R.E. of the sequential sign test relative to the sequential probability ratio (SPR) is also given by $4g^2(0)/\int (g'^2(v)/g(v))dv$. This value is also the A.R.E. of the sign test relative to the most powerful test for testing $H: \theta = \theta_0$ vs. $H_1: \theta = \theta_1$ with $\theta_1 > \theta_0$, (where observations are drawn from G with a symmetric density function $g(x-\theta)$, $x \in R$), if competing tests are considered based on a fixed sample size (cf. Hájek and Šidák (1967), Chapter 7).

6.4 Computations of the A.R.E. of certain transformations relative to the optimal procedure for some given known distributions of the error random variables:

In Table II, for the sake of easy computations (especially in the KW situation), we take the A.R.E. of a modified procedure

θ_h relative to the optimal procedure θ^* to be

$$e(\theta_h;\theta^*) = [H^2(h)/(\int (g'(v)/g(v))^2 g(v)\,dv)] .$$

TABLE II

	Value of $[H^2(h)/(\int (g'(v)/g(v))^2 g(v)\,dv)]$		A.R.E.
Transformation / Distribution of errors	$h_1(v) = v$ (RM or KW procedure)	$h_2(v) = \text{sign}(v)$ (Procedure (F))	$e(h_2;h_1)$
Normal	1	$\frac{2}{\pi}$	$\frac{2}{\pi}$
Double exponential	$\frac{1}{2}$	1	2
Uniform	Undefined	Undefined	$\frac{1}{3}$
Triangular	Undefined	Undefined	$\frac{2}{3}$
Cauchy	Undefined	$\frac{8}{\pi^2} = .810$	Undefined
Logistic	$\frac{9}{4\pi^2}$	.75	$\frac{\pi^2}{3}$

REFERENCES

REFERENCES

1. Albert, Arthur E. and Gardner, LeLand A. (1967): Stochastic Approximation and Nonlinear Regression, Research Monograph No. 42, The M.I.T. Press, Cambridge, Massachusetts.

2. Blum, J.R. (1954): Approximation Methods which Converge with Probability One, Ann. Math. Statist., 25, pp. 737-744.

3. Burkholder, D.L. (1956): On a Class of Stochastic Approximation Processes, Ann. Math. Statist., 27, pp. 1044-1059.

4. Chung, Kai Lai (1968): A Course in Probability Theory, Harcourt, Brace and World, Inc.

5. ________ (1954): On a Stochastic Approximation Method, Ann. Math. Statist., 25, 3, pp. 463-483.

6. Derman, C. and Sacks, J. (1954): On Dvoretzky's Stochastic Approximation Theorem, Ann. Math. Statist., 30, 2, pp. 601-606.

7. Dubins, Lester E. and Freedman, David (1965): A Sharper Form of the Borel-Cantelli Lemma, and the Strong Law, Ann. Math. Statist., 36, 3, pp. 800-818.

8. Dupač, V. (1957): On Kiefer-Wolfowitz Approximation Method, Casopis Pest. Mat. 82, pp. 47-75.

9. Dvoretzky, Aryeh (1956): On Stochastic Approximation, Proceedings of Third Berkeley Symposium on Math. Stat. and Probability, Vol. 1, pp. 39-59, University of California Press.

10. Fabian, V. (1960): Stochastic Approximation Methods, Czech. Math. Journal 10, pp. 123-159.

11. ________ (1964): A New One-dimensional Stochastic Approximation Method for Finding a Local Minimum of a Function, Trans. Third Prague Conf. Information Theory, Statistical Decision Functions, Random Processes, Czechoslovak Academy of Sciences, Prague, 85-105.

12. ________ (1967): Stochastic Approximation of Minima with Improved Asymptotic Speed, Ann. Math. Statist., 38, 1, pp. 191-200.

13. ________ (1968a): On the Choice of Design in Stochastic Approximation Methods, Ann. Math. Statist., 39, 2, pp. 457-465.

14. Fabian, V. (1968b): On Asymptotic Normality in Stochastic Approximation, Ann. Math. Statist., 39, 4, pp. 1327-1332.

15. ________ (1971): Stochastic Approximation, Proc. Symp. on Optimizing Methods in Statistics, Ohio State University, June 1971 (to appear). RM-280, Michigan State University.

16. Graves, L.M. (1956): The Theory of Functions of Real Variables, McGraw-Hill.

17. Groeneveld, R.A. (1971): A Note on the Sequential Sign Test, The American Statistician, 25, 2.

18. Hájek, J. and Šidák, Z. (1967): Theory of Rank Tests, Academic Press.

19. Hodges, J.L., Jr. and Lehmann, E.L. (1956): Two Approximations to the Robbins-Monro Process, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, 1, pp. 95-104.

20. Kesten, Harry (1958): Accelerated Stochastic Approximation, Ann. Math. Statist., 29, pp. 41-59.

21. Kiefer, J. and Wolfowitz, J. (1952): Stochastic Estimation of the Maximum of a Regression Function, Ann. Math. Statist., 23, pp. 462-466.

22. Loève, Michel (1963): Probability Theory, Third Edition, D. van Nostrand Company, Inc.

23. Lukacs, Eugene (1968): Stochastic Convergence, D.C. Heath and Company.

24. Pontryagin, L.S. et al (1963): The Mathematical Theory of Optimal Process, Interscience Publishers, a division of John Wiley & Sons, Inc.

25. Rao, C.R. (1965): Linear Statistical Inference and its Applications, John Wiley & Sons, Inc.

26. Robbins, H. and Monro, S. (1951): A Stochastic Approximation Method, Ann. Math. Statist., 22, pp. 400-407.

27. Rudin, Walter (1966): Real and Complex Analysis, McGraw-Hill Book Company.

28. Springer, B.G.F. (1969): Numerical Optimization in the Presence of Random Variability. The Single Factor Case, Biometrika, 56, pp. 65-74.

29. Ventor, J.H. (1967): An Extension of the Robbins-Monro Procedure, Ann. Math. Statist., 38, 1, pp. 181-190.

30. Wolfowitz, J. (1956): On Stochastic Approximation Methods, Ann. Math. Statist., 27, pp. 1151-1155.

31. Zacks, S. (1971): The Theory of Statistical Inference, John Wiley and Sons, Inc.