SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853–7501

TECHNICAL REPORT NO. 781

February 1988
Revised December 1988

EFFICIENT ESTIMATORS FROM A SLOWLY CONVERGENT ROBBINS–MONRO PROCESS

by

David Ruppert

ABSTRACT

The Robbins–Monro (RM) recursive procedure for estimating the root, θ, of an unknown regression function, M, takes the form $X_{n+1} = X_n - a_n Y_n$. Here Y_n is an unbiased (conditional upon the past) estimate of the $M(X_n)$ and $\{a_n\}$ is a positive sequence tending to 0. It is known that X_n is an asymptotically efficient estimate of θ if $a_n = 1/(\dot{M}(\theta)n)$. In an earlier paper, Frees and Ruppert showed that if $a_n = a/n$ for any a greater than $1/(2\dot{M}(\theta))$, then an asymptotically efficient estimate of θ can be obtained by fitting a least–squares line to $\{(Y_i, X_i): i = 1, \ldots, n\}$. Moreover, by choosing a large, one may obtain a more precise estimate of $\dot{M}(\theta)$ which may also be of interest.

This paper studies the RM process when $a_n = Dn^{-\alpha}$, $1/2 < \alpha < 1$ and $D > 0$. For such α the RM process differs in several interesting ways from the case $\alpha = 1$. The results of Frees and Ruppert are extended to the case $1/2 < \alpha < 1$. The estimate of $\dot{M}(\theta)$ converges to rate $O((\log n)^{-1/2})$ if $\alpha = 1$, but at rate $O(n^{(\alpha-1)/2})$ for α in $(1/2, 1)$. This suggests using $\alpha < 1$ when one is interested both in estimating θ and in estimating M in a neighborhood of θ.

Perhaps the most surprising result is that when α is in $(1/2, 1)$, then the arithmetic mean, $\bar{X} = n^{-1} \Sigma X_i$, is an asymptotically efficient estimate of θ regardless of the choice of D!

1. INTRODUCTION.

Robbins and Monro (1951) introduced the following problem. For each real x, suppose we can perform an experiment with a response, y_x, having distribution F_x. The expected response is then

$$M(x) = \int_{-\infty}^{\infty} y \, dF_x(y) .$$

In many applications, say to process control or bioassay, a real number γ is chosen and it is desired to estimate an unknown θ satisfying

$$M(\theta) = \gamma .$$

By replacing y with $(y - \gamma)$ we can assume, without loss of generality, that $\gamma = 0$.

The Robbins–Monro procedure for estimating θ lets X_1 be an arbitrary initial estimate of θ and updates by the recursion

$$X_{n+1} = X_n - a_n Y_n .$$

Here a_n is a suitable positive sequence of real numbers, and Y_n has distribution F_{X_n}. It was established by Blum (1954), that under mild conditions on M and the variance function

$$\sigma_x^2 = \int_{-\infty}^{\infty} (y - M(x))^2 \, dF_x(y) ,$$

that $X_n \to \theta$, a.s., if

(1) $$\Sigma\, a_n = \infty$$

and

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

Hanson and Goodsell (1976) further investigate consistency.

Chung (1954) showed that if $a_n = Dn^{-\alpha}$ for $D > 0$ and $1/2 < \alpha < 1$, or $D > 1/(2\dot{M}(\theta))$ and $\alpha = 1$, then

(3) $$n^{\alpha/2}(X_{n+1} - \theta) \xrightarrow{D} N(0, \sigma^2(\alpha, D)),$$

where

(4) $$\begin{aligned} \sigma^2(\alpha, D) &= D\sigma_\theta^2/(2\dot{M}(\theta)) \text{ if } 1/2 < \alpha < 1 \\ &= D^2\sigma_\theta^2/(2\dot{M}(\theta)D - 1) \text{ if } \alpha = 1\,. \end{aligned}$$

Fabian's (1968) Theorem 2.2 now provides a quicker proof of (1.3) – (1.4). These results suggest that $\alpha = 1$ is optimal, and that $D = 1/\dot{M}(\theta)$, which minimizes $\sigma^2(1, D)$, is optimal when $\alpha = 1$. Venter (1967) proposed a scheme where D is replaced by a consistent sequence of estimators, D_n, of $1/\dot{M}(\theta)$, and Lai and Robbins (1979,1981) investigate in detail methods for estimating $\dot{M}(\theta)$ so that X_n has minimal asymptotic variance. Procedures that estimate $\dot{M}(\theta)$ to achieve minimal asymptotic variance are called adaptive.

The results of Fabian (1983) (see also Fabian and Hannan (1987)) show that adaptive Robbins–Monro procedures are LAM (locally asymptotically minimax) when $F_x(\cdot) = \Phi[(\cdot - M(x)/\sigma_x)]$, Φ being the standard normal distribution. If F_x is non–Gaussian, then one can still obtain an LAM procedure by suitable transformation of the observations, Y_n. See Fabian (1973,1983). The point of Fabian's LAM results is that adaptive Robbins–Monro procedures are asymptotically efficient within the class of *all* possible estimation methods.

Because Robbins–Monro procedures use the last observation, X_{n+1}, to estimate θ, to obtain efficiency, X_n must converge to θ as rapidly as possible, and then the "design", $\{X_i: i = 1, \dots, n\}$ is highly concentrated about θ. This concentration can be a problem if one wishes to estimate $M(\cdot)$ in a neighborhood of θ, say by estimating $\dot{M}(\theta)$ and using a linear approximation. Frees and Ruppert (1987) note that the problem can be resolved by using an estimate of θ based on the entire sequence $\{(Y_i, X_i): i = 1, \dots, n\}$.

Frees and Ruppert (1987) consider the case where $a_n = D_n n^{-1}$ and $D_n \to D > 1/(2\dot{M}(\theta))$. They show that if one fits a least–squares line to $\{(Y_i, X_i): i = 1, \dots, n\}$ and lets $\hat{\theta}_n$ be the zero of this line, then

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathscr{D}} N(0, \sigma_\theta^2/(\dot{M}(\theta))^2)\,.$$

Therefore, $\hat{\theta}_n$ is asymptotically equivalent to the adaptive RM procedure, even if X_n is not efficient because $D \neq 1/\dot{M}(\theta)$. Moreover, there are potential advantages to a choice of D besides $1/\dot{M}(\theta)$. First, it may be possible to use $D_n \equiv D$ for some constant D, provided one can choose $D > 1/(2\dot{M}(\theta))$. Such procedures are easy to implement. More importantly, large values of D lead to more precise estimates of $\dot{M}(\theta)$. In many applications, e.g., bioassay, θ is a convenient location parameter describing the regression function M, but a scale parameter such as $\dot{M}(\theta)$ is also of major interest.

For estimation purposes, if $a_n = Dn^{-1}$, then the larger the value of D the better. This fact leads one to consider $a_n = Dn^{-\alpha}$, $\alpha < 1$, the topic of the present paper.

Here it is shown that if $1/2 < \alpha < 1$, then although X_n converges to θ only at rate $n^{-\alpha/2}$, there exist two simple, asymptotically efficient estimators of θ. The first is $\bar{X}_{n+1}$ where

$$\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i ,$$

and the second is the least–squares estimator proposed by Frees and Ruppert (1987). Thus, we extend the Frees and Ruppert results from $\alpha = 1$ to $1/2 < \alpha < 1$. The discovery that $\bar{X}_n$ is efficient was quite surprising to me, although I was not aware of related work of Bather (1988) at that time.

Another reason for using $a < 1$ is to increase the rate at which the large deviation probability

$$P(|X_n - \theta| > c)$$

converges to 0 for fixed, positive c; see sections 4 and 6.

In section 2, notation and assumptions are presented. Section 3 contains representation theorems that elucidate the structure of the processes $\{X_n\}$, $\{\bar{X}_n\}$, and the sequence of least–squares estimators. This section also gives the basic results on asymptotic distributions. Section 4 contains a simulation study. Section 5 contains the proofs and several technical lemmas. Section 6 is a discussion and summary.

2. NOTATION, DEFINITIONS, AND ASSUMPTIONS.

All random variables are defined on a probability space $(\Omega, \mathscr{F}, P)$. All relations between random variables are meant to hold with probability 1. $[x]$ is the greatest integer less than or equal to x. "$O(\cdot)$" and "$o(\cdot)$" notation have their usual meaning, and we write $X_n \sim Y_n$ if $X_n/Y_n \to 1$. We write $X_n \xrightarrow{\mathscr{D}} X$ if X_n converges in distribution to X.

ASSUMPTION 2.1: Let D be a positive number, let α be in $(1/2, 1)$, let M map $\mathbb{R}^1$ to $\mathbb{R}^1$, let $\{\mathscr{F}_n : n \geq 0\}$ be an increasing sequence of σ–sub–algebras of $\mathscr{F}$, let X_1 be an $\mathscr{F}_0$ measurable random variable, and for each $n \geq 1$ let

(1) $$X_{n+1} = X_n - Dn^{-\alpha}\{M(X_n) + \epsilon_n\}$$

ASSUMPTION 2.2: Let

(1) $$E^{\mathscr{F}_{n-1}}\epsilon_n = 0 \text{ for all } n \geq 1 ,$$

let

(2) $$\mathrm{Var}^{\mathscr{F}_{n-1}}\epsilon_n \to \sigma^2 > 0 \text{ as } n \to \infty ,$$

let $\delta > \alpha^{-1} - 1$, and let

(3) $$\sup_n E^{\mathscr{F}_{n-1}}|\epsilon_n|^{4(1+\delta)} < \infty .$$

ASSUMPTION 2.3: Let θ be the unique solution to

(1) $$M(\theta) = 0 ,$$

and let M have two continuous derivatives in a neighborhood of θ. Let $\dot{M}(\theta)$ be positive.

Assumption 2.4: Let $X_n \to \theta$.

Definitions 2.5: Define $Y_n = M(X_n) + \epsilon_n$,

$$\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i , \quad \bar{Y}_n = n^{-1} \sum_{i=1}^{n} Y_i ,$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i (X_i - \bar{X}_n)}{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2} ,$$

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n ,$$

and

$$\hat{\theta}_n = - \hat{\beta}_0 / \hat{\beta}_1 .$$

Remarks 2.6: Equation (2.1.1) can be written $X_{n+1} = X_n - Dn^{-\alpha} Y_n$. $\hat{\beta}_1$ and $\hat{\beta}_0$ are, respectively, the least–squares estimates of slope and intercept when Y is regressed on X using a straight–line model. $\hat{\theta}_n$ is the root of the least–squares line. The smoothness assumption 2.3 and the consistency assumption 2.4, make the use of the straight–line model reasonable. See Wu (1985) and Frees and Ruppert (1987) for further discussion. Assumption 2.4 follows from standard results, e.g., Theorem 1 of Robbins and Siegmund (1971) (see their application 4) and additional, mild assumptions on M.

DEFINITIONS 2.7: Let $\beta_1 = \dot{M}(\theta)$, and let $\beta_0 = -\beta_1 \theta$.

REMARK 2.8: $\ell(x) = \beta_1(x - \theta) = \beta_0 + \beta_1 x$ is the tangent line to M at $(\theta, M(\theta))$.

DEFINITION 2.9: For positive integers i and $k > i$ define

$$(1) \qquad c(i, k) = \exp\left(-D\beta_1 \sum_{\ell=i+1}^{k} \ell^{-\alpha}\right).$$

3. MAIN RESULTS.

THEOREM 3.1: Assume 2.1, 2.3, and 2.4. Then

$$(1) \qquad X_{n+1} - \theta = B_{n+1}^{-1}\{B_{n_0}(X_{n_0} - \theta) - \sum_{i=n_0}^{n} Di^{-\alpha} B_{i+1}\epsilon_i\}$$

where

$$(2) \qquad B_{j+1} = \prod_{i=n_0}^{j} \{1 - Di^{-\alpha}\dot{M}(\delta_i)\}^{-1}$$

for a sequence δ_i such that

$$(3) \qquad |\delta_i - \theta| < |X_i - \theta|$$

and

$$(4) \qquad n_0 = \inf\{n: Di^{-\alpha}\dot{M}(\delta_i) < 1 \text{ for all } i \geq n\}.$$

COROLLARY 3.2: Assume 2.1, 2.2, 2.3, and 2.4. Then

(1) $$X_{n+1} - \theta = -D \sum_{i=1}^{n} c(i, n)\, i^{-\alpha} \epsilon_i + o(1) .$$

Moreover,

(2) $$X_{n+1} - \theta = -Dn^{-\alpha} \sum_{i=N(n)}^{n} c(i, n) \epsilon_i + o(1) ,$$

where $N(n) = [n - Kn^{\alpha} \log n]$ and K is a sufficiently large positive constant.

THEOREM 3.3: Assume 2.1, 2.2, 2.3, and 2.4. Then

(1) $$n^{1/2} (\bar{X}_{n+1} - \theta) = -1/(\beta_1 n^{1/2}) \sum_{i=1}^{n} \epsilon_i + o(1) ,$$

(2) $$n^{1/2} (\hat{\theta}_n - \theta) = -1/(\beta_1, n^{1/2}) \sum_{i=1}^{n} \epsilon_i + O(n^{(\alpha-1)/2}(\log_2 n)) ,$$

and

(3) $$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n} \epsilon_i (X_i - \theta)}{\sum_{i=1}^{n} (X_i - \theta)^2} + O(n^{-\alpha/2}(\log n)^{3/2}) .$$

COROLLARY 3.4: Assume 2.1, 2.2, 2.3, and 2.4. Then, letting $\sigma^2 = \sigma^2_{\theta}$,

(1) $$n^{1/2}(\bar{X}_{n+1} - \theta) \xrightarrow{\mathscr{D}} N(0, \sigma^2/\beta_1^2)\,,$$

(2) $$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathscr{D}} N(0, \sigma^2/\beta_1^2)\,,$$

and

(3) $$n^{(1-\alpha)/2}(\hat{\beta}_1 - \beta_1) \xrightarrow{\mathscr{D}} N(0, 2(1-\alpha)/D)\,.$$

DISCUSSION 3.5: Equations (3.1.1) – (3.1.4) hold for $\alpha = 1$ and have been used by many authors beginning with Sacks (1958). However, B_{n+1} behaves differently in the case $\alpha = 1$ compared to $\alpha < 1$, and in the former case instead of (3.2.1) holding, Ruppert (1982, Theorem 4.2) has shown that

(1) $$X_{n+1} - \theta = -n^{-1} D \sum_{i=1}^{n} (i/n)^{D\beta_1 - 1} \epsilon_i + o(1)\,.$$

The above expression shows that $(X_{n+1} - \theta)$ is essentially a weighted average of $\epsilon_1, \dots, \epsilon_n$. In contrast, when $\alpha < 1$ equation (3.2.2) shows that $(X_{n+1} - \theta)$ is essentially a weighted average of only the last $(Kn^{\alpha} \log n)$ of the ϵ_i. The asymptotic distributions given in (1.3) – (1.4) can be easily derived from (1) and (3.2.2). If $0 < p_1 < p_2 < 1$, then (3.2.2) shows that $X_{[np_1]}$ and $X_{[np_1]}$ are asymptotically uncorrelated if $\alpha \in (\frac{1}{2}, 1)$, but (1) shows these to be asymptotically correlated if $\alpha = 1$.

Result (3.4.2) holds when $\alpha = 1$ provided that $D > 1/(2\beta_1)$ (Frees and Ruppert (1987)), but (3.4.1) will not hold if $\alpha = 1$. In fact, letting $\Delta = D\beta_1 - 1$, (1) implies that

(2) $$\bar{X}_n - \theta = -n^{-1}D \sum_{k=1}^{n} k^{-1} \sum_{i=1}^{k} (i/k)^{\Delta}\epsilon_i + o(1)$$

$$= -n^{-1}D \sum_{i=1}^{n} i^{\Delta} \left(\sum_{k=1}^{n} k^{-1-\Delta}\right)\epsilon_i + o(1)$$

$$\sim n^{-1} \left(\frac{D}{\Delta}\right) \sum_{i=1}^{n} \{(i/n)^{\Delta} - 1\}\, \epsilon_i + o(1)\,.$$

Since (3.4.1) and (3.4.2) hold for all $D > 0$ and α in $(1/2, 1)$, (3.4.3) suggests taking α close to $1/2$ and D large. Clearly, further research is necessary to guide the choice of α and D in practical situations where n is finite. Some work along this direction appears in the next section. If $\alpha = 1$, then

$$(\log n)^{1/2} (\hat{\beta}_1 - \beta_1) \xrightarrow{\mathscr{D}} N(0, (2a-1)/(a\beta_1)^2)$$

(Frees and Ruppert (1987), equation (2.2)), which when contrasted with (3.4.3) shows the potential of using $\alpha < 1$.

4. MONTE CARLO

A small simulation study was performed using the regression function

(1) $$M(x) = \frac{\kappa}{1 + e^{-x/\kappa}} - \frac{\kappa}{2}$$

The results reported here are for $\kappa = 3$. The algorithm (2.1.1) was replaced by

$$X_2 = X_1 - .5, \text{ and}$$

$$X_{n+1} = X_n - \left[\frac{D}{(\hat{\beta}_n \vee \delta) n^{\alpha}} Y_n \right]_{-\Delta}^{\Delta}, \quad n \geq 2,$$

where $\delta = .01$, $\Delta = 1$, and

$$[\mu]_{-1}^{1} = (-\Delta \vee \mu) \wedge \Delta.$$

Here $\hat{\beta}_n$ is the least–squares slope estimator:

$$\hat{\beta}_n = \frac{\Sigma_{i=1}^{n} Y_i (X_i - \bar{X}_n)}{\Sigma_{i=1}^{n} (X_i - \bar{X}_n)^2} .$$

Experimentation showed that truncating $\hat{\beta}_n$ below by δ and truncating the step size at $\pm \Delta$ resulted in an algorithm that was much less variable than the untruncated version.

Except for the truncation points, δ and Δ, the algorithm is scale equivariant, i.e., equivariant to the transformation $Y \to bY$, $b \neq 0$. Without scale invariance, the algorithm's performance would depend crucially on the product $(D\dot{M}(\theta))$, and a value of D that worked well in the simulations for a particular M could not be recommended for other M.

The conditional distribution of Y_n, given the past, was normal with mean $M(X_n)$ and standard deviation $\frac{1}{4}$. The number of observations was $N = 10, 40$, or 250. There were 500 simulations for $N = 10$ and 250 for the other sample sizes.

The parameters D and α were varied as shown in Table 1. Two values, .75 and 1.5, were used for X_1, but they produced similar results, so only $X_1 = 1.5$ was reported in Table 1. That table contains the root mean square errors (RMSE) for three estimators of θ: X_{N+1} (RM), $\bar{X}_{N+1}$ ($\bar{X}$), and $\hat{\beta}_N$ (LS). In the computation of $\bar{X}$, the first two X's, X_1 and $X_2 = X_1 - .5$, were excluded. The following conclusions can be reached from examination of Table 1:

(1) When $N = 10$ or 40, LS with $\alpha = .6$ and $D = 1$ or 1.5 is superior to best RM estimator.

(2) When $N = 40$ or 250, then RM with $\alpha = .6$ is less efficient than $\alpha = 1$. This agrees with asymptotics. However, when $\alpha = 1$, then RM with $D = 1.5$ is slightly *more* efficient than $D = 1$, in disagreement with asymptotics but similar to the findings of Frees and Ruppert (1987).

(3) When $N = 250$, the best LS and best RM estimators are roughly comparable, LS being only slightly more efficient.

(4) As predicted by asymptotics, $\bar{X}$ with $\alpha = 1$ is inefficient, but $\bar{X}$ with $\alpha = .6$ and $D = 1.5$ is an excellent estimator, comparable to the best RM and LS estimators.

Squared bias is a very small proportion, often less than one–hundredth, of the mean square errors in Table 1. For this reason, bias was not reported.

Table 2 reports the standard deviation and bias of $\hat{\beta}_N$ as an estimator of $\dot{M}(\theta)$. Typically, $\hat{\beta}_N$ is positively biased. This cannot be due to the nonlinearity of M, since $\dot{M}$ reaches its maximum at θ so the bias due to nonlinearity is downward. Because X_n is a function of $e_1,\ldots,e_{n-1}$, $\hat{\beta}_N$ is biased even if M is linear; see Walters (1985) who discovered this bias in control problems similar to stochastic approximation.

Increasing the sum of squares, $\Sigma_{i=1}^{N} (X_i - \bar{X})^2$, by using $\alpha = .6$ and/or $D = 1.5$ and/or $X_1 = 1.5$ tends to decrease both the standard deviation and bias of $\hat{\beta}_N$, especially for $N = 40$ or 250. (Note that if the bias were due to nonlinearity then we would expect the bias to increase with the sum of squares.) It is interesting that if $X_1 = .75$, then one needs almost 250 observations to achieve the same accuracy as $n = 10$ and $X_1 = 1.5$.

To increase nonlinearity, κ in (1) was changed from 3 to 1. The extra nonlinearity increased the variability of all three estimators of θ, but did not substantially change their relative efficiencies. The effect on $\hat{\beta}_n$ was to *decrease* both variance and bias, especially bias.

Large deviations

Another potential use of $\alpha < 1$ is for control problems where one needs to keep X_n close to θ for all n. Lai and Robbins (1979,1981) suggest the control loss

$$\sum_{i=1}^{n} (X_i - \theta)^2 \tag{2}$$

for which $\alpha = 1$ and $D = 1$ is asymptotically optimal, but in many situations a large deviation of X_n from θ will be of particular concern. For example, an excessive drug dosage may cause death, while a slightly suboptional dosage may have no serious consequences. In such situations, the loss is better measured by the rate at which

$$P(|X_n - \theta| > c) \tag{3}$$

(or perhaps $P(X_n - \theta > c)$) converges to 0 for some fixed constant c.

Table 3 reports Monte Carlo estimates of (3) for $\alpha = .6$ and 1, $c = .1, .2, .4, .6,$ and .8, $D = 1$, $n = 5, 10,$ and 20, and $X_1 = .75$ and 3.0. All estimates are based on 2000 simulations.

When $n = 5$, then X_n is more concentrated about θ when $\alpha = .6$ than when $\alpha = 1$. When $n = 10$, then $\alpha = .6$ and $\alpha = 1$ produce comparable results for $X_1 = .75$, but when $X_1 = 3$ then X_n is more concentrated around θ for $\alpha = .6$.

When $X_1 = 3$, then

$$P(|X_5 - \theta| > .8)$$

is much smaller for $\alpha = .6$ than $\alpha = 1$.

In summary, X_n approaches a fixed neighborhood, say $[\theta - c, \ \theta + c]$, more rapidly for $\alpha = .6$ than $\alpha = 1$, and this effect is most pronounced when X_1 is far from θ and c is large. This finding agrees with theoretical results of Berger (1978); see section 6.

5. PROOFS AND TECHNICAL COMPLEMENTS.

5.1 **Proof of Theorem 3.1:** By (2.11) and assumptions 2.3 and 2.4, there exist δ_i such that $|\delta_i - \theta| < |X_i - \theta|$ and $X_{n+1} = X_n - Dn^{-\alpha}(\dot{M}(\delta_n)X_n + \epsilon_n)$, so that

$$B_{n+1}X_{n+1} = B_nX_n - B_{n+1}Dn^{-\alpha}\epsilon_n \, . \tag{1}$$

Iterating (1) back to n_0, (3.1.1) is proved. □

Lemma 5.2: Assume 2.1, 2.2, 2.3, and 2.4. Then

$$\limsup_{n\to\infty} n^{\alpha/2}|X_n - \theta| \ / \ [(1-\alpha)\log n]^{1/2} = (D/\beta_1)^{1/2}\sigma \, . \tag{1}$$

Proof: By Theorem 3.1, the lemma is a special case of the following lemma with $H_i = D$, $v_i = \epsilon_i$, $\eta = \sigma$, $\Delta = -\alpha$, and $c = D\beta_1$. The extra generality of Lemma 5.3 will be used later. To verify (5.3.3), use assumption (2.4) to show that

$$\log B_{j+1} \sim D \sum_{\ell=n_0}^{j} \ell^{-\alpha}\,\dot{M}(\delta_\ell) \sim D\beta_1 \frac{j^{1-\alpha}}{1-\alpha} \, . \ \square$$

Lemma 5.3: Let α be in $(1/2, 1)$. Let $\{\mathcal{G}_n\}$ be an increasing sequence of σ–algebras, and let $(v_i, \mathcal{G}_i)$ be a martingale difference sequence such that

(1) $$\sup_n E(|v_n|^{2(1+\alpha)} | \mathscr{G}_{n-1}) < \infty$$

for some $\delta > \alpha^{-1} - 1$, and

(2) $$E(v_n^2 \mid \mathscr{G}_{n-1}) \to \eta^2 > 0 .$$

Let L_i and B_{i+1} be $\mathscr{G}_{i-1}$ measurable, let $L_i \to L$ for a positive constant L, let $c > 0$, and let

(3) $$\log B_{i+1} \sim ci^{1-\alpha}/(1-\alpha) .$$

Finally, let $Y_n = L_n n^{\Delta} B_{n+1} v_n$ for some real Δ, and define

$$W_n = \sum_{i=1}^{n} Y_i .$$

Then

(4) $$\limsup_{n\to\infty} \frac{|W_n|}{\{n^{\Delta+\alpha/2} B_{n+1}[(1-\alpha)\log n]^{1/2}\}} = \frac{L\eta}{c^{1/2}} .$$

PROOF: Define

$$s_n^2 = \sum_{i=1}^{n} E[Y_i^2 \mid \mathscr{G}_{i-1}] \quad \text{and} \quad u_n^2 = 2\log_2 s_n^2 .$$

By Theorem 3 of Stout (1970),

(5)
$$\limsup_{n\to\infty} \frac{|W_n|}{s_n u_n} = 1$$

if

(6)
$$\sum_{n=1}^{\infty} (K_n s_n)^{-2} u_n^2 \, E\,\{Y_n^2 \, I[Y_n^2 > s_n^2 K_n^2 / u_n^2] \mid \mathcal{G}_{n+1}\} < \infty$$

for a sequence $\{K_n\}$ such that K_n is $\mathcal{G}_{n-1}$ measurable and $K_n \to 0$. We will use $K_n = (\log n)^{-1/2}$. Let $m(n) = [n - n^{\alpha} (\log n)^2]$ and $D = c/(1-\alpha)$. Since

$$s_n^2 / (L^2 \eta^2) \sim (\sum_{i=1}^{n} i^{2\Delta} B_{i+1}^2) \sim B_{n+1}^2 (\sum_{i=m(n)}^{n} i^{2\Delta} (B_{i+1} / B_{n+1})^2)$$

$$\sim B_{n+1}^2 \, n^{2\Delta} \sum_{i=m(n)}^{n} \ell^{-2Dn^{-\alpha}(n-i)} ,$$

it follows that, $s_n^2 \sim B_{n+1}^2 \, n^{2\Delta+\alpha} \{L^2 \eta^2 / (2D)\}$ and $u_n^2 \sim 2 \log_2 B_{n+1} \sim 2\,(1-\alpha) \log n$. Next, choose $\delta > \alpha^{-1} - 1$ so that (1) holds. Then

$$E\,\{Y_n^2 \, I\,[Y_n^2 > s_n^2 K_n^2 / u_n^2] \mid \mathcal{G}_{n-1}\} \le E\,\{Y_n^{2(1+\delta)} (u_n^2 / s_n^2 K_n^2)^{\delta} \mid \mathcal{G}_{n-1}\}$$

$$= O\,\{n^{2\Delta - \alpha\delta} \, B_{n+1}^2 (\log n)^{2\delta}\} ,$$

and

$$(K_n s_n)^{-2} u_n^2 = O((\log n)^2 \, B_{n+1}^{-2} \, n^{-(2\Delta+\alpha)})$$

so that (6) holds since $\delta > \alpha^{-1} - 1$. Moreover,

(7) $$s_n u_n \,/\, \{n^{\Delta+\alpha/2} B_{n+1}[2\,(1-\alpha)\log n]^{1/2}\} \to H\eta \,/\, (2D)^{1/2}$$

as $n \to \infty$, and (4) follows from (5) and (7). □

Lemma 5.4: Assume 2.1, 2.2, 2.3, and 2.4. Then (3.3.1) holds and

(1) $$\limsup_{n\to\infty} \frac{n^{1/2}(\bar{X}_n - \theta)}{(2\,\log_2 n)^{1/2}} = \frac{\sigma}{\beta_1}.$$

Proof: (1) follows from (2.2.2), (3.3.1), and the LIL for martingales, so it suffices to prove (3.3.1).

We first note that, without loss of generality, we can assume that M is linear. To see this, note that if $M^*(x) = \dot{M}(\theta)(X-\theta)$, $X_0^* = X_0$, and

$$X_{n+1}^* = X_n^* - Dn^{-\alpha}\{M^*(X_n^*) + \epsilon_n\},$$

then X_n^* satisfies the hypothesis of Lemma 5.2 so (5.2.1) holds with X_n replaced by X_n^*. Therefore, there exists a positive constant such that

$$|X_{n+1} - X_{n+1}^*| \le (1 - Dn^{-\alpha}\dot{M}(\theta))\,|X_n - X_n^*| + Kn^{-2\alpha}\log n$$

for all large n. Then by Chung's Lemma (Fabian (1971), Lemma 3.1),

(2) $$\limsup_{n\to\infty} \frac{n^{\alpha}\,|X_n - X_n^*|}{\log n} < \infty.$$

Define $\bar{X}_n^* = n^{-1} \sum_{i=1}^{n} X_i^*$. By (2)

$$n^{1/2}|\bar{X}_n - \bar{X}_n^*| = O(n^{-1/2} \log n \sum_{i=1}^{n} i^{-\alpha}) = O(n^{-1/2} \log n) = o(1) ,$$

so if (3.3.1) holds for $\bar{X}_n^*$, then (3.3.1) also holds for $\bar{X}_n$.

We now proceed under the assumption that M is linear. By (3.3.1)

$$\sum_{k=n_o}^{n} (X_{k+1} - \theta) = \{\sum_{k=n_o}^{n} B_{k+1}^{-1}\} B_{n_o}(X_{n_o} - \theta) - D \sum_{i=n_o}^{n} i^{-\alpha} (\sum_{k=i}^{n} \frac{B_{i+1}}{B_{k+1}})\epsilon_i . \tag{3}$$

Since M is linear,

$$\log B_{j+1} = D\beta_1 \sum_{\ell=n_o}^{j} \ell^{-\alpha} + O(\sum_{\ell=n_o}^{j} \ell^{-2\alpha}) . \tag{4}$$

Therefore, if $(k-j) > j^{\alpha} (\log j)^2$, then

$$\frac{B_{j+1}}{B_{k+1}} = o(j^{-(\log j)D\beta_1/2}) \text{ as } j \to \infty , \tag{5}$$

and if $0 < (k-j) \le j^{\alpha} (\log j)^2$, then

$$\frac{B_{j+1}}{B_{k+1}} = c(j,k) \{1 + O(j^{-2\alpha} (k-j))\} . \tag{6}$$

Next, using the notation $q(i) = [i + i^{\alpha} (\log i)^2]$

$$(7) \qquad \left| \sum_{i=n_o}^{n} i^{-\alpha} \sum_{k=i}^{n} \left[\frac{B_{i+1}}{B_{k+1}} - c(i,k)\right] \epsilon_i \right|$$

$$\leq \sum_{i=n_o}^{n} i^{-\alpha} \left[\sum_{k=i}^{q(i)} O(i^{-2\alpha}(k-i))\right] |\epsilon_i| + O(1)$$

$$= O\left(\sum_{i=n_o}^{n} i^{-3\alpha}(q(i)-i)^2\right) + O(1) = o(n^{1/2})$$

since $\alpha > 1/2$. It follows from (5) and (6) that

$$(8) \qquad \sum_{k=n_o}^{\infty} B_{k+1}^{-1} \text{ converges.}$$

Also, by (2.9.1)

$$(9) \qquad \sum_{i=n_o}^{n} i^{-\alpha} \left\{ \sum_{k=i}^{n} c(i,k) \right\} \epsilon_i \sim \sum_{i=n_o}^{n} i^{-\alpha} \left\{ \sum_{k=1}^{\min(n,q(i))} c(i,k) \right\} \epsilon_i$$

$$\sim \sum_{i=n_o}^{n} i^{-\alpha} \left[\int_i^{\min(n,q(i))} \exp(-D\beta_1 i^{-\alpha}(x-i))dx \right] \epsilon_i \sim (D\beta_1)^{-1} \sum_{i=n_o}^{n} \epsilon_i \, .$$

By (3), (7), (8), and (9) it follows that

$$n^{-1/2} \sum_{k=n_o}^{n} (X_{k+1} - \theta) = -\frac{1}{\beta_1 \sqrt{n}} \sum_{i=1}^{n} \epsilon_i + o(1) \, ,$$

which proves (3.3.1) . □

5.5 **Proof of Corollary 3.2:** (3.2.1) and (3.2.2) follow from (3.1.1), (5.4.4), and (5.4.5). □

Lemma 5.6: Assume 2.1, 2.2, 2.3, and 2.4. Then

$$(1) \qquad \sum_{j=1}^{n} (X_j - \theta) \sim \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 \sim \frac{D\sigma^2 n^{1-\alpha}}{2(1-\alpha)} .$$

Proof: Define

$$H_n = \prod_{i=1}^{n} [\max\{(1 - Di^{-\alpha}\dot{M}(\delta_i)), 1/2\}]^{-1}$$

where δ_i is given by Theorem 3.1. From (3.1.1) – (3.1.2) it follows that

$$(2) \qquad X_{n+1} = H_n^{-1}\{\rho_o - \sum_{i=1}^{n} DH_i i^{-\alpha}\epsilon_i\}$$

for some random variable ρ_o. Let $C = D/(1-\alpha)$, $\beta = (1-\alpha)$, and

$$(3) \qquad \tau_n = H_n \exp(-Cn^\beta) .$$

For n sufficiently large

$$\tau_n - \tau_{n-1} = \left\{[1 - Dn^{-\alpha}\dot{M}(\delta_n)]^{-1}\exp[-C(n^\beta - (n-1)^\beta)] - 1\right\} \cdot H_{n-1}\exp(-C(n-1)^\beta)$$

$$= \left\{[1 - Dn^{-\alpha}\dot{M}(\delta_n)]^{-1}[1 - C\beta n^{\beta-1} + O(n^{\beta-2})] - 1\right\} \tau_{n-1}$$

so that

$$(4) \qquad \tau_n - \tau_{n-1} = O(n^{-2\alpha}\tau_n) .$$

Let

$$\tilde{S}_j = \rho_o - \sum_{i=1}^{j} D\tau_i \exp(Ci^{\beta})\, i^{-\alpha}\epsilon_i .$$

It follows from (2) that

$$(5) \qquad \sum_{j=2}^{n+1} X_j^2 = \sum_{j=1}^{n} \tau_j^{-2} \exp(-2Cj^{\beta})\tilde{S}_j^2 .$$

Define

$$(6) \qquad a(j) = \sum_{k=j}^{\infty} \exp(-2Ck^{\beta}) .$$

Following Lai and Robbins (1979, proof of Theorem 4(i)), we have

$$(7) \qquad \sum_{j=1}^{n} \tau_j^{-2} \exp(-2Cj^{\beta})\tilde{S}_j^2 = \sum_{j=1}^{n} \tau_j^{-2}\{a(j) - a(j+1)\}\, \tilde{S}_j^2$$

$$= \sum_{j=2}^{n} a(j)\{\tau_j^{-2} - \tau_{j-1}^{-2}\}\, \tilde{S}_j^2 + \sum_{j=2}^{n} a(j)\, \tau_{j-1}^{-2}(\tilde{S}_j^2 - \tilde{S}_{j-1}^2)$$

$$- a(n+1)\tau_n^{-2}\, \tilde{S}_n^2 + a(1)\tau_1^{-2}\, \tilde{S}_1^2 = Q_1 + Q_2 + Q_3 + Q_4 \text{, say.}$$

From $\tau_j^{-2} - \tau_{j-1}^{-2} = (\tau_{j-1} - \tau_j)(\tau_j + \tau_{j-1})\tau_j^{-2}\,\tau_{j-1}^{-2}$ and (4) it follows that

(8) $$\sum_{j=2}^{n} a(j)\,|\,\tau_j^{-2} - \tau_{j-1}^{-2}\,|\,\tilde{S}_j^2 = o(\sum_{j=2}^{n} a(j)\, j^{-2\alpha}\, \tau_j^{-2\alpha}\, \tilde{S}_j^2) + O(1)\,.$$

Next

(9) $$a(j) = \exp(-2Cj^{\beta}) \sum_{k=1}^{\infty} \exp(-2C(k^{\beta} - j^{\beta}))$$

$$\sim \exp(-2Cj^{\beta}) j^{\alpha} / (2D)\,,$$

so that by (5) and (8)

(10) $$Q_1 = o(\sum_{j=2}^{n+1} X_j^2) + O(1)\,.$$

By Lemma 5.3 with $\Delta = -\alpha$, $B_{i+1} = H_i$, $L_i = 1$, and $v_n = \epsilon_n$, it follows that

(11) $$\tilde{S}_n = O(H_n n^{-\alpha/2} (\log n)^{1/2})\,,$$

whence

(12) $$Q_3 = O(\exp(-2Cn^{\beta}) n^{\alpha}\, \tau_n^{-2}\, \tilde{S}_n^2) = O(H_n^{-2} n^{\alpha}\, \tilde{S}_n^2) = O(\log n)\,.$$

Now

$$Q_2 = \sum_{j=2}^{n} a(j)\, \tau_{j-1}^{-2}\, D^2 \tau_j^2 \exp(2Cj^{\beta}) j^{-2\alpha} \epsilon_j^2$$

$$-2 \sum_{j=2}^{n} a(j)\, \tau_{j-1}^{-2}\, \tilde{S}_{j-1} D \tau_j \exp(Cj^{\beta}) j^{-\alpha} \epsilon_j = T_1 + T_2\text{, say.}$$

By (9)

(13) $$T_1 \sim \frac{D}{2} \sum_{j=2}^{n} j^{-\alpha} \epsilon_j^2 \sim \frac{D\sigma^2}{2(1-\alpha)} n^{1-\alpha},$$

by the martingale convergence theorem and (2.2.2). By (9) and (11)

$$T_2 = O\{\sum_{j=2}^{n} \tau_{j-1}^{-2}\, \tau_j \exp(-Cj^{\beta}) j^{-\alpha/2} (\log j)^{1/2} H_j G_j \epsilon_j\}$$

for a sequence of $\{G_n\}$ such that G_n is F_{n-1} measurable and $G_n = O(1)$. Therefore,

(14) $$T_2 = O[(\sum_{j=2}^{n} j^{-\alpha} \log j)^{1/2} (\log_2 n)^{1/2}] = o(n^{(1-\alpha)/2} (\log n))$$

by a law of the iterated logarithm for martingales, e.g., Stout (1970, Theorem 3). By (5), (7), (10), (12), (13), and (14)

$$\sum_{j=1}^{n} X_{j+1}^2 = o(\sum_{j=1}^{n} X_{j+1}^2) + O(n^{(1-\alpha)/2} \log n) + \frac{D\sigma^2}{2(1-\alpha)} n^{1-\alpha},$$

so that

(15) $$\sum_{j=1}^{n} X_{j+1}^2 \sim \frac{D\sigma^2}{2(1-\alpha)} n^{1-\alpha}.$$

The lemma follows from (15) and (5.4.1). □

5.7 **Proof of Theorem 3.3:** (3.3.1) was proved in Lemma 5.4

Define $\bar{\epsilon}_n = n^{-1} \sum_{i=1}^{n} \epsilon_i$. Then by assumption (2.3) and (4.6.1), we have

(1) $$\bar{Y}_n = \beta_1(\bar{X}_n - \theta) + \bar{\epsilon}_n + O(n^{-\alpha}) .$$

Next

(2) $$\sum_{i=1}^{n} Y_i(X_i - \bar{X}_n) = \beta_1 \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

$$+ (1/2) \sum_{i=1}^{n} M^{(2)}(\eta_i)(X_i - \bar{X}_n)(X_i - \theta)^2 + \sum_{i=1}^{n} (X_i - \bar{X}_n)\epsilon_i$$

where $\eta_i \to \theta$. By (5.2.1) and (5.4.1)

(3) $$\sum_{i=1}^{n} |X_i - \bar{X}_n|(X_i - \theta)^2 = O(n^{-3\alpha/2+1} (\log n)^{3/2}) .$$

Then (2), (3), (3.3.1), and (5.6.1) imply (3.3.3). From (3.3.3) and the law of the iterated logarithm for martingales,

(4) $$\hat{\beta}_1 - \beta_1 = O(n^{(\alpha-1)/2} (\log_2 n)^{1/2}) + O(n^{-\alpha/2} (\log n)^{3/2})$$

$$= O(n^{(\alpha-1)/2} (\log_2 n)^{1/2}) ,$$

since $-\alpha < -1/2 < \alpha - 1$. By (1)

$$(\hat{\theta} - \theta) = \left(\frac{-(\bar{Y}_n - \hat{\beta}_1 \bar{X}_n)}{\hat{\beta}_1} - \theta\right)$$

$$= (\bar{X}_n - \theta)\left(\frac{\hat{\beta}_1 - \beta_1}{\hat{\beta}_1}\right) + \frac{\bar{\epsilon}_n}{\hat{\beta}_1} + O(n^{-\alpha})$$

and (3.3.2) follows from (4) and (5.4.1). □

5.8 **Proof of Corollary 3.4:** The corollary is consequence of the CLT for martingales, e.g., Corollary 3.1 of Hall and Heyde (1980). Only the proof of (3.4.3) is nontrivial. Define

$$X_{ni} = \frac{\epsilon_i (X_i - \theta)\, n^{(1-\alpha)/2}}{\sum_{i=1}^{n} (X_i - \theta)^2} .$$

By (2.2.2) and (5.6.1)

$$\text{(1)} \qquad \sum_{i=1}^{n} E\,[X_{ni}^2 \mid \mathcal{F}_{i-1}] \sim \frac{\sigma^2\, n^{1-\alpha}}{\sum_{i=1}^{n} (X_i - \theta)^2} \sim \frac{2(1-\alpha)}{D} .$$

By (2.2.3), (5.4.1), and (5.6.1)

$$\text{(2)} \qquad \sum_{i=1}^{n} E[X_{ni}^2\, I(|X_{ni}| > \epsilon) \mid \mathcal{F}_{i-1}] \le \epsilon^{-1} \sum_{i=1}^{n} E[|X_{ni}|^3 \mid \mathcal{F}_{i-1}] = o(1) .$$

By (1) and (2), the assumptions of Hall and Heyde's Corollary 3.1 hold with $\eta^2 = 2(1-\alpha)/D$, $\mathcal{F}_{n,i} = \mathcal{F}_i$, and $k_n = n$. □

6. DISCUSSION AND SUMMARY

This paper examines the Robbins–Monro procedure when the "tuning constants", $\{a_n\}$, converge to 0 at rate $n^{-\alpha}$, $1/2 < \alpha < 1$. It is well–known that for such a_n, X_n is *not* asymptotically efficient. However, we have found that an asymptotically efficient estimate can be constructed by fitting a least–squares line to all the data. Also the sample mean of $X_1,\ldots,X_{n+1}$ is asymptotically efficient.

There are two advantages of using $\alpha < 1$: (1) the least–squares estimate of $\dot{M}(\theta)$ is improved and, (2) the rate at which the large–deviation probability, $P(|X_n - \theta| > c)$, $c > 0$, converges is improved, at least under some circumstances.

Berger's (1978) large–deviation theorem gives a theoretical underpinning to (2). Suppose that $P(|\epsilon_n| \le L) = 1$ for some $L < \infty$ and M is monotonic with $M(\infty) = \sup_x M(x)$.

Suppose $\alpha = 1$. If $M(\infty) < L$, then by Komlo's and Révész (1972) for all $c > 0$ and $\delta > 0$

$$P(X_n - \theta > c) \le \exp(-n^{M(\infty)/L-\delta})$$

for all large n. This rate can, of course, be arbitrarily slow if $M(\infty)/L$ is small enough.

Now suppose $1/2 < \alpha < 1$. Then for all c there exists $\eta > 0$ such that

$$P(X_n - \theta > c) \le \exp(-\eta n^{2\alpha-1})$$

for all large n (Berger, 1978, Theorem 3.1).

Recently, Bather (1988) has studied a procedure that, in our notation, can be written as

$$X_{n+1} = \bar{X}_n - n a_n \bar{Y}_n.$$

His heuristic argument suggests that $\bar{X}_n$ is asymptotic optimal if $a_n = an^{-\alpha}$, $0 < \alpha < 1$. Bather advocates these estimators because of their simplicity. $\bar{X}_n$ from an ordinary (nonadaptive) RM process with $1/2 < \alpha < 1$ can be recommended for the same reason.

ACKNOWLEDGMENTS: I thank Vaclav Fabian for his valuable comments and encouragement, and especially for suggesting the problem of large deviations and supplying the reference to Berger (1978). I thank John Bather for helpful comments and for kindly providing a preprint of his interesting paper.

REFERENCES

Bather, J.A. (1988). *Stochastic approximation: a generalization of the Robbins–Monro procedure.* In **Proceedings of the Fourth Prague Symposium on Asymptotic Statistics** (to appear).

Berger, Erich (1978). *Einige ergebnisse über das asymptotische verhalten der verteilungen bei stochastischen approximationverfahren.* Diplomarbeit, Georg–August–Universität, Gottingen.

Chung, K.L. (1954). *On a stochastic approximation method.* **Annals of Mathematical Statistics** 25, 463–483.

Fabian, V. (1968). *On asymptotic normality in stochastic approximation.* **Annals of Mathematical Statistics** 39, 1327–1332.

Fabian, V. (1973). *Asymptotically efficient stochastic approximation: the RM case.* **Annals of Statistics** 1, 486–495.

Fabian, V. (1983). *A local asymptotic minimax optimality of an adaptive Robbins–Monro stochastic approximation procedure.* **Mathematical Learning Models – Theory and Algorithms** (U. Herkenrath, D. Kalin, and W. Vogel, eds.), **Lecture Notes in Statistics** 20, Springer–Verlag: New York.

Fabian, V. and Hannan, J. (1987). *Local asymptotic behavior of densities.* **Statistics and Decisions** 5, 105–138.

Frees, E.W. and Ruppert, D. (1987). *Estimation following a Robbins–Monro designed experiment.* **Technical Report** 811, Department of Statistics, University of Wisconsin.

Goodsell, C.A. and Hanson, D.L. (1976). *Almost sure convergence for the Robbins–Monro process.* **Annals of Probability** 4, 890–891.

Hall, P. and Heyde, C. (1980). **Martingale Limit Theory and Its Applications.** Academic Press: New York.

Komlós, J. and Révész, P. (1972). *On the rate of convergence of the Robbins–Monro method.* **Z. Wahrscheinlichkeitstheorie und verw. Gebiete** 25, 39–47.

Lai, T.L. and Robbins, H. (1979). *Adaptive designs and stochastic approximation.* **Annals of Statistics** 7, 1196–1221.

Lai, T.L. and Robbins, H. (1981). *Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes.* **Z. Wahrscheinlichkeitstheorie Verw. Gebiete** 56, 329–366.

Robbins, H. and Siegmund, D. (1971). *A convergence theorem for non–negative almost supermartingales and some applications.* **Optimizing Methods in Statistics** (J.S. Rustagi, ed.). Academic Press: New York, 233–257.

Sacks, J. (1958). *Asymptotic distribution of stochastic approximation procedures.* **Annals of Mathematical Statistics** 29, 373–405.

Stout, W. (1970). *A martingale analogue of Kolmogorov's law of the iterated logarithm.* **Z. Wahrscheinlichkeitstheorie Verw. Gebiete** 15, 279–290.

Venter, J. (1967). *An extension of the Robbins–Monro procedure.* **Annals of Mathematical Statistics** 38, 181–190.

Walters, C.J. (1985). *Bias in the estimation of functional relationships from time series data.* **Can. J. Fish. Aquat. Sci** 42, 147–149.

Wu, C. (1985). *Efficient sequential designs with binary data.* **Journal of the American Statistics Association** 80, 974–984.

Table 1: RMSE of three estimators of θ

α	D	Method	N = 10	N = 40	N = 250
.6	1	RM	.463	.265	.138
		$\bar{X}$	.393	.185	.065
		LS	.406	.167	.063
	1.5	RM	.534	.323	.169
		$\bar{X}$	.369	.173	.064
		LS	.390	.166	.063
1	1	RM	.437	.200	.067
		$\bar{X}$	.488	.287	.100
		LS	.528	.182	.067
	1.5	RM	.426	.181	.067
		$\bar{X}$	.422	.229	.079
		LS	.436	.169	.065
Fisher information bound			.316	.158	.063

Table 2: Standard deviation and bias of $\hat{\beta}_N$

α	D		N = 10		N = 40		N = 250	
			X_0=.75	X_0=1.5	X_0=.75	X_0=1.5	X_0=.75	X_0=1.5
.6	1	sd	.158	.105	.139	.084	.093	.069
		bias	.108	.039	.086	.032	.066	.040
	1.5	sd	.154	.102	.130	.082	.077	.060
		bias	.100	.038	.073	.030	.047	.031
1	1	sd	.154	.110	.143	.093	.123	.081
		bias	.106	.037	.105	.039	.111	.053
	1.5	sd	.160	.106	.154	.089	.121	.079
		bias	.111	.039	.112	.038	.111	.053

Table 3: Estimates of large deviations probabilities

$X_1 = .75$

c	n=5 α=.6	n=5 α=1	n=10 α=.6	n=10 α=1	n=20 α=.6	n=20 α=1
.1	.860	.890	.812	.838	.710	.750
.2	.699	.776	.618	.663	.475	.516
.4	.436	.522	.313	.358	.172	.177
.6	.247	.299	.135	.161	.058	.044
.8	.124	.158	.066	.063	.020	.010

$X_1 = 3$

c	n=5 α=.6	n=5 α=1	n=10 α=.6	n=10 α=1	n=20 α=.6	n=20 α=1
.1	.924	.978	.826	.900	.754	.790
.2	.849	.954	.671	.794	.529	.584
.4	.688	.892	.374	.565	.225	.253
.6	.483	.786	.186	.356	.074	.080
.8	.318	.658	.090	.172	.021	.012