Statistica Sinica **25** (2015), 1679-1689
doi:http://dx.doi.org/10.5705/ss.2014.197

A SKEWED VERSION OF THE ROBBINS-MONRO-JOSEPH PROCEDURE FOR BINARY RESPONSE

Dianpeng Wang[1], Yubin Tian[1] and C. F. Jeff Wu[2]

[1]*Beijing Institute of Technology and* [2]*Georgia Institute of Technology*

Abstract: The Robbins-Monro stochastic approximation procedure has been used for sensitivity testing. Joseph (2004) recognized that it is not well suited for binary data and proposed a modification that gives better performance for p between 0.1 and 0.9. However, for extreme p values, say $p \leq 0.01$ or $p \geq 0.99$, the Joseph version does not perform well. To overcome this difficulty, we propose a modification based on an asymmetric quadratic loss function. The new procedure can speed up convergence by employing different penalties for undershooting and overshooting to reduce the expected loss. Simulation comparisons show the clear advantages of the new procedure for extreme p values.

Key words and phrases: Asymmetric loss function, bioassay, extreme quantile, sensitivity testing, stochastic approximation.

1. Introduction

Sensitivity testing has been used in bioassays and industries. Their applications include finding the dose level at which 50% of the specimens survive (Wetherill (1963); Wu (1985)), finding the level of shock necessary to make 99.99% of the explosives to fire (Neyer (1994)), and finding a stack force level at which the problem of multi-feeds in a paper feeder occurs with 0.1% probability (Joseph and Wu (2002)). Primary interest lies in finding the stimulus level at which a given response (or success) occurs with probability p. The stimulus level can be the dose of drug or the applied voltage, and the response can be survival or explosion, etc. Let x denote the stimulus level, y the binary response, success or failure, and $F(x)$ the probability of success at level x. The problem can be formulated as that of finding the p^{th} quantile ζ_p such that $F(\zeta_p) = p$. For this problem, many procedures have been proposed in the literature, e.g., Robbins and Monro (1951), Wu (1985), Neyer (1994), Joseph, Wu, and Tian (2007), Dror and Steinberg (2008) and Wu and Tian (2014). Joseph (2004) proposed a modification to the Robbins-Monro procedure in the case of binary data, and showed that it gives better performance for p between 0.1 and 0.9. However, for extreme quantiles like $p \leq 0.01$ or $p \geq 0.99$, its performance can suddenly deteriorate, as

demonstrated in the simulation study of Wu and Tian (2014). The main prognosis in Wu and Tian is that the iterative sequence of the Joseph procedure can descend extremely slowly toward a target like $\zeta_{0.99}$ if the starting value of the iteration is guessed far above the target. To mitigate this problem, we propose to use an asymmetric loss function around the target to penalize above target more severely than below target. The minimization of the expected loss will force the iterative sequence to make bigger descending steps.

This paper is organized as follows. In Section 2, the Robbins-Monro procedure and the Joseph modification are reviewed. In Section 3, the proposed procedure is developed. In Section 4, simulation comparisons between the new procedure and the Joseph modification are given. Concluding remarks are given in Section 5.

2. Robbins-Monro Procedure and Joseph's Modification

Robbins and Monro (1951) proposed the stochastic approximation procedure

$$x_{n+1} = x_n - a_n(y_n - p), \tag{2.1}$$

where y_n is the response at the stress level x_n, $\{a_n\}$ is a sequence of positive constants, and p is pre-specified by the experimenter. Robbins and Monro (1951) suggested choosing $a_n = c/n$, where c is a constant. Chung (1954), Hodges and Lehmann (1956), and Sacks (1958) recommended the choice $c = \{F'(\zeta_p)\}^{-1}$, where F' denotes the derivative of F. Since $F'(\zeta_p)$ is unknown, an adaptive choice is $c = 1/\hat{\beta}_n$, where $\hat{\beta}_n$ is the regression slope of y_i over x_i. Anbar (1973) and Lai and Robbins (1981) proved that $\hat{\beta}_n$ with proper truncation converges to $F'(\zeta_p)$ and the procedure with the truncated $\hat{\beta}_n$ has the same asymptotic distribution as the procedure with $c = \{F'(\zeta_p)\}^{-1}$. Wetherill (1963), Cochran and Davis (1965), and Young and Easterling (1994) studied the performance of the Robbins-Monro procedure and several modifications through simulations and concluded that the Robbins-Monro procedure and related modifications should be used only for p in the immediate neighborhood of 0.5.

Recognizing that, for binary response y, the Robbins-Monro procedure does not make provision for the binary nature of data, Joseph (2004) proposed the modification

$$x_{n+1} = x_n - a_n(y_n - b_n), \tag{2.2}$$

where $\{a_n\}$ and $\{b_n\}$ are sequences of constants. Start the iteration at some level x_1, believed to be close to $\theta = \zeta_p$, and assume the prior distribution of θ has $E(\Theta) = x_1$ and $\text{var}(\Theta) = \tau_1^2$. Joseph proposed to choose a_n and b_n by minimizing $E(z_{n+1}^2)$ subject to the condition $E(z_{n+1}) = 0$, where $z_n = x_n - \Theta$. The solution

to this minimization problem is obtained by choosing

$$b_n = \Phi\{\frac{\Phi^{-1}(p)}{(1+\beta^2\tau_n^2)^{1/2}}\}, \quad a_n = \frac{1}{b_n(1-b_n)}\frac{\beta\tau_n^2}{(1+\beta^2\tau_n^2)^{1/2}}\phi\{\frac{\Phi^{-1}(p)}{(1+\beta^2\tau_n^2)^{1/2}}\},$$
$$\tau_{n+1}^2 = \tau_n^2 - b_n(1-b_n)a_n^2, \quad \beta = \frac{F'(F^{-1}(p))}{\phi(\Phi^{-1}(p))}\frac{1}{\sigma}, \tag{2.3}$$

where $\Phi(\cdot)$ is the normal distribution function and $\phi(\cdot)$ is the normal density function. For brevity, we refer to (2.2)-(2.3) as the Robbins-Monro-Joseph procedure.

A simulation study in Joseph (2004) showed the superior performance of the modification over the original Robbins-Monro procedure. However, these simulations were performed only at different values of p between 0.1 and 0.9. Wu and Tian (2014) performed simulations at $p = 0.9$, $p = 0.99$ and $p = 0.999$ and showed that the performance of the Robbins-Monro-Joseph procedure depends critically on the starting value x_1, especially for $p = 0.99$ and $p = 0.999$. It can perform badly when x_1 is far from the target value ζ_p. A typical example of this poor performance is given in (Wu and Tian, 2014, p.11). For estimating $\zeta_{0.99}$ (= 12.3263), suppose x_1 is chosen to be 19.3054. In the next 60 iterations the corresponding y_i are all equal to 1 and thus the corresponding x_i sequence makes small, decreasing steps. Even the terminating x value x_{61}(= 17.2733) is still much larger than 12.3263. In this situation, the sequence descends very slowly toward ζ_p, which explains why the Robbins-Monro-Joseph procedure has much higher mean square errors than competing procedure in the simulations studied by Wu and Tian (2014). In order to accelerate its convergence toward ζ_p, we need to modify the procedure so that bigger steps can be made and more $y_i = 0$ observed. One way to achieve this is to use an asymmetric loss function around the target ζ_p to penalize a positive deviation of $x_i - \zeta_p$ more severely than a negative deviation. This drives the x_i sequence to make bigger descending steps, especially for extreme values of p. The detailed development is given in the next section. The asymmetric loss function has been used effectively before in deriving an optimization procedure in robust parameter design (Moorhead and Wu (1998)).

3. The New Procedure

Consider the procedure in (2.2) with the starting level x_1. Assume the distribution function $F(x)$ is from the location-scale family $F(x) = G((x-\mu)/\sigma)$, where G is a known distribution function. Let $\theta = \zeta_p$ and assume that the prior distribution for θ has $E(\Theta) = \xi$ and var(Θ)= τ_1^2. Let $z_n = x_n - \Theta$. Then (2.2) is

$$z_{n+1} = z_n - a_n(y_n - b_n). \tag{3.1}$$

Denote $P(y_n = 1 \mid z_n)$ by $M(z_n)$, so $M(z_n) = G(z_n/\sigma + G^{-1}(p))$. Consider the asymmetric quadratic loss function

$$L(z) = wz^2, \tag{3.2}$$

with $w = \lambda_1$ for $z \le 0$ and λ_2 for $z > 0$, where $\lambda_1 > 0$ and $\lambda_2 > 0$. We propose to choose x_1 to minimize $E\{L(z_1)\}$ and a_n and b_n to minimize $E\{L(z_{n+1})\}$. Let $\upsilon_n = E(z_n)$ and $\tau_n^2 = \text{var}(z_n)$. From (3.1), we have

$$\upsilon_{n+1} = E(z_{n+1}) = \upsilon_n - a_n E\{M(z_n)\} + a_n b_n, \tag{3.3}$$

$$\tau_{n+1}^2 = a_n^2[E\{M(z_n)\} - E^2\{M(z_n)\}] - 2a_n[E\{z_n M(z_n)\} - \upsilon_n E\{M(z_n)\}] + \tau_n^2. \tag{3.4}$$

Clearly, it is difficult to obtain an exact distribution of z_n or an exact formula for $E\{L(z_n)\}$. We employ an approach similar to that in Joseph (2004, p.463) to approximate the distribution of z_n by $N(\upsilon_n, \tau_n^2)$ and $M(z_n)$ by $\Phi\{\Phi^{-1}(p) + \gamma z_n\}$, where

$$\gamma = \frac{M'(0)}{\phi\{\Phi^{-1}(p)\}}. \tag{3.5}$$

The choice of γ in (3.5) is to guarantee that the approximation and M have the same slope at 0. With these approximations, we have

$$\begin{aligned} E\{L(z_{n+1})\} = {} & (\lambda_1 - \lambda_2)\Big\{(\tau_{n+1}^2 + \upsilon_{n+1}^2)\Phi(-\frac{\upsilon_{n+1}}{\tau_{n+1}}) - \upsilon_{n+1}\tau_{n+1}\phi(-\frac{\upsilon_{n+1}}{\tau_{n+1}})\Big\} \\ & + \lambda_2(\tau_{n+1}^2 + \upsilon_{n+1}^2). \end{aligned} \tag{3.6}$$

Choose υ_1 to minimize $E\{L(z_1)\}$ for a given τ_1 and let $x_1 = \xi + \upsilon_1$. Let $\lambda = \lambda_1/\lambda_2$ and call it the *skewness coefficient*. By minimizing (3.6) with respect to a_n and υ_{n+1}, we obtain the optimal a_n and υ_{n+1} as the roots of the equations:

$$\frac{\partial E\{L(z_{n+1})\}}{\partial a_n} = \lambda_2 \tau'_{n+1}[2\lambda\Phi(-\frac{\upsilon_{n+1}}{\tau_{n+1}})\tau_{n+1} + 2\tau_{n+1}\{1 - \Phi(-\frac{\upsilon_{n+1}}{\tau_{n+1}})\}] = 0, \tag{3.7}$$

$$\frac{\partial E\{L(z_{n+1})\}}{\partial \upsilon_{n+1}} = 2\lambda_2(\lambda - 1)\Big\{\upsilon_{n+1}\Phi(-\frac{\upsilon_{n+1}}{\tau_{n+1}}) - \tau_{n+1}\phi(-\frac{\upsilon_{n+1}}{\tau_{n+1}})\Big\} + 2\lambda_2\upsilon_{n+1} = 0, \tag{3.8}$$

where υ_{n+1} and τ_{n+1} are defined in (3.3) and (3.4). That the solution to (3.7) $-$(3.8) is indeed a minimum is proven in Proposition 1. It is easy to show from (3.7) that

$$a_n = \frac{E\{z_n M(z_n)\} - \upsilon_n E\{M(z_n)\}}{E\{M(z_n)\} - E^2\{M(z_n)\}}. \tag{3.9}$$

Using numerical computation, we can obtain the root υ_{n+1} of (3.8). From (3.3), we can obtain the optimal b_n as

$$b_n = E\{M(z_n)\} - \frac{(\upsilon_n - \upsilon_{n+1})}{a_n}. \tag{3.10}$$

The proposed skewed version of the Robbins-Monro-Joseph procedure is defined by (3.1), (3.9) and (3.10). At the nth iteration, its step length is the absolute value of $a_n(y_n - b_n)$. Since a_n is given and b_n is close p (as proved in Theorem 1), the step length is determined by $|y_n - p|$. For p close to 1, it is much smaller when $y_n = 1$ than when $y_n = 0$. Thus, in the specific example given after (2.3), all the 60 iterations make small descending steps toward the target when only $y_n = 1$ is observed. By choosing the skewness coefficient λ to be smaller than 1, the skewed procedure forces the iterations to make bigger downward steps than upward steps, which also increases the chance of observing $y_n = 0$. Similarly, for p close to 0, we should choose λ larger than 1 to enforce upward bias. Thus, on the right panel of Figure 1 in Section 4 we choose large values of λ^{-1}, while on the left panel we choose large values of λ. A further study in Section 4 shows how this modification can help mitigate the problem encountered in this specific example.

Proposition 1. *The values of a_n and b_n that minimize $E\{L(z_{n+1})\}$ in (3.6) are unique.*

Proof. Since $\lambda > 0$, $\tau_{n+1} > 0$, and $0 < \Phi(-\upsilon_{n+1}/\tau_{n+1}) < 1$, the sign of $\partial E\{L(z_{n+1})\}/\partial a_n$ in (3.7) is the same as that of τ'_{n+1}. By taking the derivative of τ^2_{n+1} in (3.4), we have

$$2\tau_{n+1}\tau'_{n+1} = 2a_n[E\{M(z_n)\} - E^2\{M(z_n)\}] - 2[E\{z_nM(z_n)\} - \upsilon_nE\{M(z_n)\}]. \tag{3.11}$$

Because $E\{M(Z_n)\} < 1$, the expression on the right of (3.11) is strictly linear in a_n and thus (3.7) has a unique root. Because $\lim_{\upsilon_{n+1}\to-\infty} \partial E\{L(z_{n+1})\}/\partial\upsilon_{n+1} = -\infty$, $\lim_{\upsilon_{n+1}\to\infty} \partial E\{L(z_{n+1})\}/\partial\upsilon_{n+1} = \infty$, and $\partial^2 E\{L(z_{n+1})\}/\partial\upsilon^2_{n+1} = 2(\lambda - 1)\lambda_2\Phi(-\upsilon_{n+1}/\tau_{n+1}) + 2\lambda_2 \geq 0$, (3.8) has a unique root. A direct computation has the determinant of the Hessian matrix as $4[E\{M(z_n)\} - E^2\{M(z_n)\}][(\lambda - 1)\lambda_2\Phi(-\upsilon_{n+1}/\tau_{n+1}) + \lambda_2]^2$, which is larger than 0. Therefore, the values a_n and b_n that minimize $E\{L(z_{n+1})\}$ in (3.6) are unique.

Because the $E\{M(z_n)\}$ and $E\{z_nM(z_n)\}$ terms in (3.9) and (3.10) do not have exact formulas, we resort to the normal approximations

$$\begin{aligned} E\{M(z_n)\} &\approx \Phi\Big(\frac{\Phi^{-1}(p) + \gamma\upsilon_n}{\sqrt{1+\gamma^2\tau_n^2}}\Big), \\ E\{z_nM(z_n)\} &\approx \frac{\gamma\tau_n^2}{\sqrt{1+\gamma^2\tau_n^2}}\phi\Big(\frac{\Phi^{-1}(p) + \gamma\upsilon_n}{\sqrt{1+\gamma^2\tau_n^2}}\Big) + \upsilon_n\Phi\Big(\frac{\Phi^{-1}(p) + \gamma\upsilon_n}{\sqrt{1+\gamma^2\tau_n^2}}\Big), \end{aligned} \tag{3.12}$$

where γ is given in (3.5). If $\lambda > 1$, $\lambda_1 > \lambda_2$, then the penalty for under-shooting is more severe than for over-shooting. Therefore, in order to reduce the expected loss $E\{L(z_n)\}$, the corresponding υ_n value should be positive. Similarly, for $\lambda < 1$, $\upsilon_n < 0$ and for $\lambda = 1$, $\upsilon_n = 0$.

Table 1. Average and RMSE of x_{61} for estimating $\zeta_{0.99}(= 12.3263)$, $n = 60$, true model=normal distribution, $\mu_g = 10$, $\sigma_g = 4.0$, and $x_1 = 19.3054$.

λ^{-1}	Average	RMSE	$\%\{y_i = 0\}$
1	17.2733	4.9470	0
100	15.3308	3.0047	0.3%
1,000	14.6634	2.3658	37.4%

Proposition 2. *If $\lambda = 1$, then $\upsilon_n \equiv 0$; if $\lambda > 1$, then $\upsilon_n > 0$; if $\lambda < 1$, then $\upsilon_n < 0$.*

Proof. If $\lambda = 1$, the solution of (3.8) is $\upsilon_{n+1} = 0$. If $\lambda > 1$, because $\partial E\{L(z_{n+1})\}/\partial\upsilon_{n+1} < 0$ at $\upsilon_{n+1} = 0$ and because $\partial^2 E\{L(z_{n+1})\}/\partial\upsilon_{n+1}^2 \geq 0$, the root υ_{n+1} of (3.8) is positive. A similar argument can be made for $\lambda < 1$.

Theorem 1. *For the procedure given in* (3.1), (3.9), *and* (3.10) *with the normal approximation* (3.12), $z_n \to^p 0$.

The proof is in the Appendix.

4. Simulation Comparisons

We first show how the modified procedure can mitigate the problem in the example of Section 3. The example came from Table 6A in Wu and Tian (2014) with the choice of $\mu_g = 10$ and $\sigma_g = 4$, where μ_g is the same as the mean of the normal distribution and σ_g is the same as γ^{-1} with γ defined in (3.5). Using the parameters of Wu and Tian, we generated 1,000 simulations for the modified procedure with $\lambda^{-1} = 100$ and 1,000. In Table 1, the average value and root mean square error (RMSE) of x_{61} and the percentage of $y_i = 0$ among the 1,000 simulations are given for each of the three competing versions. The values in the Table 4.1 for $\lambda = 1$ were taken from Table 6A of Wu and Tian. Here, as λ^{-1} increases the RMSE decreases, the final estimate x_{61} gets closer to the target, and the percentage of $y_i = 0$ increases, clearly demonstrating the effectiveness of using the skewness coefficient λ. We carried out a more comprehensive simulation study to compare the performance of the procedure given in (3.1), (3.9), (3.10) with that of the Robbins-Monro-Joseph procedure. Three distributions were considered: the normal, $F(x) = \Phi(x)$; the logistic, $F(x) = (1 + e^{-1.8138x})^{-1}$; the extreme value, $F(x) = exp\{-exp(-x)\}$.

We chose $\sigma = (1.8138)^{-1}$ for the logistic distribution so that the normal and the logistic distributions have the same standard deviation; this can be important for comparing the performances under the two distributions (Wu and Tian, 2014). As in Joseph (2004), we let $\tau_1 = 1$, $\xi = F^{-1}(p)$ and chose 20 samples to estimate θ, and assumed that $M'(0) > 0$ known. We used the final value of the iterations, x_{21}, as the estimate of θ for both procedures. For

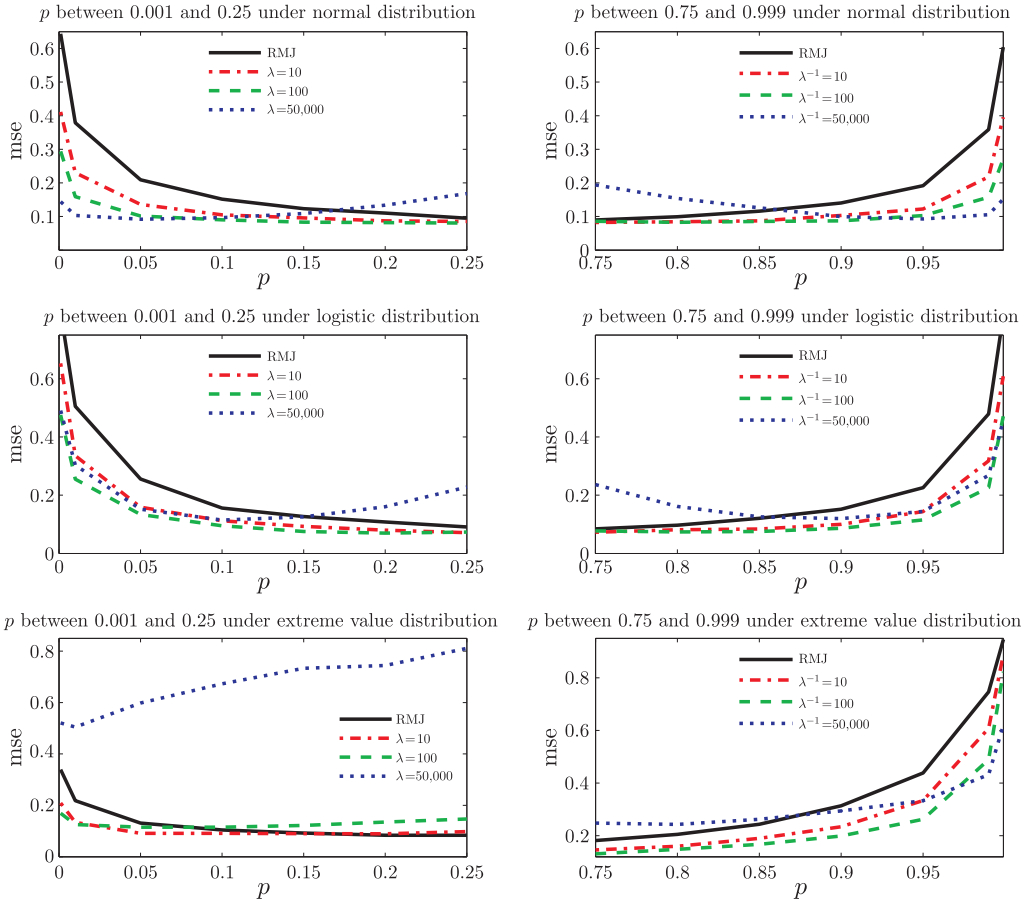

Figure 1. Plot of mean square errors of x_{21} for various procedures.

the Robbins-Monro-Joseph procedure, the starting level was randomly generated from $N(\xi, \tau_1^2)$ (Joseph (2004)). For the new procedure the starting level was randomly generated from $N(\xi + \upsilon_1, \tau_1^2)$, where υ_1 was chosen by minimizing $E\{L(z_1)\}$. For p between 0.001 and 0.25, we chose λ = 3.5, 10, 50, 100, 1,000, 50,000 in the simulation study; for p between 0.75 and 0.999, we chose λ^{-1} = 3.5, 10, 50, 100, 1,000, 50,000. For each p, 1,000 simulations were performed. And the mean square errors (MSEs) of the estimate were calculated. The results are shown in Figure 1, where the Robbins-Monro-Joseph procedure is denoted by RMJ and the proposed procedure with λ = 10, 100, 50,000 or λ^{-1} = 10, 100, 50,000 are considered. The new procedure with $\lambda = 10, 100$ or $\lambda^{-1} = 10, 100$ performs better than the Robbins-Monro-Joseph procedure, except for the $\lambda = 100$ for moderate p values under the extreme value distribution. Results for λ (or λ^{-1}) = 3.5, 50, 1,000 are omitted in order to keep the clarity of the graphs; these results do not change the overall conclusions. The same simulation study was

performed for the skewed logistic distribution $F(x) = (1+e^{-x})^{-2}$ and the cauchy distribution $F(x) = \arctan(x)/\pi + 1/2$. Because the results were similar to those for the logistic distribution, they are omitted here. As λ increases, although the new procedure is slightly worse for moderate p values, it still maintains a clear advantage over the Robbins-Monro-Joseph procedure for extreme p values. (These results are given in tables in the supplement.) However when the skewness coefficient λ takes extremely large values, like $\lambda = 50{,}000$ in Figure 1, the new procedure can be undesirable, especially under the extreme value distribution.

5. Conclusions and Further Remarks

We have propose a skewed version of the Robbins-Monro-Joseph procedure for binary data. Simulation study has shown that, for estimating the extreme quantiles, it outperforms the original procedure over a wide range of values for the skewness coefficient λ unless λ or λ^{-1} is exceedingly large. Our simulations do not tell us exactly what value of λ should be chosen. The optimal choice of λ is a complicated function of various factors, and theory does not help here. Simulations do suggest that a wide range of λ values give good performance. To find a good λ value requires some tweaking on the part of the experimenter.

An advantage of the proposed procedure is its flexibility in allocating more runs above or below target by adjusting the value of λ. This can be more easily seen for p values closer to 0.5. Consider the problem of estimating the effective dose of a compound, where is common to use a sequential design procedure for estimating or approximating the median dose $\zeta_{0.5}$; this is the ED50 problem in bioassay (Finney (1978)). In this context, the Robbins-Monro-Joseph procedure generates roughly equal numbers of $y = 1$ and $y = 0$. If $y = 1$ means death or damage of an experimental subject, it is desirable to reduce the number of $y = 1$. This can be easily achieved by using the proposed procedure with an appropriate value of λ larger than 1. If λ is moderately larger than 1, then the loss of estimation efficiency relative to the Robbins-Monro-Joseph procedure is small. A small simulation study was done to compare the skewed version (with $\lambda = 10$) and the original version (with $\lambda = 1$) in terms of the conflicting goals of estimation efficiency and reducing numbers of $y = 1$. In Table 2, the results of MSE and $\sharp$ of $\{y_i = 1\}, i = 1, \ldots, 20$ are reported for the normal, logistic and extreme value distributions. The loss of efficiency by using $\lambda = 10$ ranges from mild (normal) to moderate (logistic and extreme value) while the reduction of $\sharp$ of $\{y_i = 1\}$ is consistently around 30%. Of course a moderately large λ does not have the flexibility in allocating the sample sizes between above target and below target. The reverse is true if λ is chosen to be much larger than 1. How to make a proper choice between the two objectives is a subject for further study.

Table 2. MSE and average of $\sharp\{y_i = 1\}$ of estimating ED50 with 20 sample for three distributions.

	MSE		$\sharp\{y_i = 1\}$	
	$\lambda = 1$	$\lambda = 10$	$\lambda = 1$	$\lambda = 10$
Normal distribution	0.0791	0.0817	10.008	7.093
Logistic distribution	0.0723	0.1149	10.072	7.291
Extreme value distribution	0.1073	0.1424	10.054	7.296

For the sake of clarity, we have focused our discussions on the extreme high quantiles like $p \geq 0.99$. Extreme low quantiles are equally important in practice. For example, measures such as LD_1, the dosage required to kill 1% of the test population, are useful for the safety study of compounds. Our discussions and conclusions can be similarly made by changing from p to $1 - p$.

Acknowledgements

The co-author Wu would like to dedicate this paper to the celebration of Professor George Tiao's 80th birthday and to record his appreciation and deep gratitude to George for his tutelage and friendship over more than 35 years. The authors would like to thank the referees for their valuable comments. The research of Wang and Tian is supported by NSFC grant 11371054. The research of Wu is supported by ARO grant W911NF-14-1-0024.

Appendix

Joseph (2004) gave the following result about the convergence of the procedure in (3.1).

Lemma A.1. If (1) $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$ and (2) $\sum_{n=2}^{\infty} a_n |b_n - p| \sum_{j=1}^{n-1} a_n < \infty$ hold, then $z_n \to^p 0$.

Here we show that the sequences $\{a_n\}$ and $\{b_n\}$ given in (3.9) and (3.10) with approximate formula (3.12) satisfy the conditions of Lemma A1 and thus prove Theorem 1.

Proof. Let $\eta_n = \gamma^2 \tau_n^2$. From (3.4), (3.9), and (3.12), we have

$$\eta_{n+1} = \eta_n - \frac{\eta_n^2 I\{(\Phi^{-1}(p) - \gamma \upsilon_n)/\sqrt{1 + \eta_n}\}}{(1 + \eta_n)}, \tag{A.1}$$

where $I(u) = \phi^2(u)/[\Phi(u)\{1 - \Phi(u)\}]$ is the Fisher information of u in binary data with probability of response equal to $\Phi(u)$. From the proof of Proposition 1 in Joseph (2004), we have $\eta_n \to 0$, or equivalently $\tau_n^2 \to 0$, which together with (3.8) implies $\upsilon_n \to 0$. Thus $z_n \to^p 0$ if F is the normal distribution function.

For general F, let $h(\eta_n)$ be the right side of (A.1), $n^* = 1/I(2\Phi^{-1}(p))$ and $\eta^* = 1/I^2(2\Phi^{-1}(p))$. As in the proof of Proposition 2 in Joseph (2004), we can show that there exists an $\tilde{n}$ and some η' and η'' such that $\eta_n \le \eta'/n$ and $\eta_n \ge \eta''/n$ for all $n \ge \tilde{n}$, and that $a_n \le \eta'/(\sqrt{2\pi}p(1-p)n\gamma)$ and $a_n \ge 4\eta''\phi(\Phi^{-1}(p))/\{\gamma(n+1)\}$ for all $n \ge \tilde{n}$. Thus $\{a_n\}$ satisfies condition (1) of Lemma A1.

Next we verify condition (2). From (3.8) we have $\upsilon_{n+1}/\tau_{n+1} = (\lambda-1)\phi(-\upsilon_{n+1}/\tau_{n+1})/\{(\lambda-1)\Phi(-\upsilon_{n+1}/\tau_{n+1})+1\}$. Let $g(-\upsilon_{n+1}/\tau_{n+1})$ be the right side of this equation. Because $g(-x)-x=0$ has a unique root, there exists an M such that $\upsilon_{n+1} = M\tau_{n+1}$. From (3.10), we have $a_n(b_n-p) = a_n[\Phi\{(\Phi^{-1}(p)+\gamma\upsilon_n)/(1+\gamma^2\tau_n^2)^{1/2}\}-p] - M(\tau_n-\tau_{n+1})$. Using Taylor series expansion we obtain $\Phi\{(\Phi^{-1}(p)+\gamma\upsilon_n)/(1+\gamma^2\tau_n^2)^{1/2}\} = p + M\gamma\phi\{\Phi^{-1}(p)\}\tau_n + O(\tau_n^2)$. From (3.4), we have $\tau_n - \tau_{n+1} = a_n^2\Phi\{(\Phi^{-1}(p)+\gamma\upsilon_n)/(1+\eta_n)^{1/2}\}[1-\Phi\{(\Phi^{-1}(p)+\gamma\upsilon_n)/(1+\eta_n)^{1/2}\}]/(\tau_n+\tau_{n+1})$. Since $\eta_n = \gamma^2\tau_n^2$ was previously shown to be of the order $1/n$, $\tau_n = O(n^{-1/2})$ and also $a_n = O(1/n)$. Thus we obtain $a_n|b_n-p| = O(n^{-3/2})$. This implies that condition (2) of Lemma A1 holds. Thus by Lemma A1, $z_n \to^p 0$.

References

Anbar, D. (1973). On optimal estimation methods using stochastic approximation procedure. *Ann. Statist.* **1**, 1175-1184.

Chung, K. L. (1954). On a stochastic approximation method. *Ann. Math. Statist.* **25**, 463-483.

Cochran, W. G. and Davis, M. (1965). The Robbins-Monro method for estimating the median lethal dose. *J. Roy. Statist. Soc. Ser. B* **27**, 28-44.

Dror, H. A. and Steinberg, D. M. (2008). Sequential experimental designs for generalized linear models. *J. Amer. Statist. Assoc.* **103**, 288-297.

Finney, D. J. (1978). *Statistical Method in Biological Assay.* Oxford University Press, London.

Hodges, J. L. and Lehmann, E. L. (1956). Two approximation to the Robbins-Monro process. *Proc. Third Berkeley Symposium* **1**, 39-55. University of California, Berkeley.

Joseph, V. R. (2004). Efficient Robbins-Monro procedure for binary data. *Biometrika* **91**, 461-470.

Joseph, V. R. and Wu, C. F. J. (2002). Operating window experiments: A novel approach to quality improvement. *J. Quality Tech.* **34**, 345-354.

Joseph, V. R., Wu, C. F. J. and Tian, Y. B. (2007). Adaptive designs for stochastic root-finding. *Statist. Sinica* **17**, 1549-1565.

Lai, T. L. and Robbins, H. (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Z. Wahrsch. Verw. Gebiete* **56**, 329-360.

Moorhead, P. R. and Wu, C. F. J. (1998). Cost-driven parameter design. *Tchnometrics* **40**, 111-119.

Neyer, B. T. (1994). D-optimality-based sensitivity test. *Technometrics* **36**, 61-70.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **29**, 373-405.

Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29**, 373-405.

Wetherill, G. B. (1963). Sequential estimation of quantal response curves. *J. Roy. Statist. Soc. Ser. B* **25**, 1-48.

Wu, C. F. J. (1985). Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* **80**, 974-984.

Wu, C. F. J. and Tian, Y. B. (2014) Three-phase optimal design of sensitivity experiments (with discussions). *J. Statist. Plann. Inference* **149**, 1-32.

Young, L. J. and Easterling, R. G. (1994). Estimation of extreme quantiles based on sensitivity tests: A comparative study. *Technometrics* **36**, 48-60.

Department of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081, China.

E-mail: wdp345@gmail.com

Department of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081, China.

E-mail: tianyb@bit.edu.cn

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA.

E-mail: jeffwu@isye.gatech.edu

(Received June 2014; accepted November 2014)