

An Adequate Choice of Initial Sample Size for Selection Approach

Mohammad H. Almomani, and Rosmanjawati Abdul Rahman

Abstract—In this paper, we consider the effect of the initial sample size on the performance of a sequential approach that used in selecting a good enough simulated system, when the number of alternatives is very large. We implement a sequential approach on $M/M/1$ queuing system under some parameter settings, with a different choice of the initial sample sizes to explore the impacts on the performance of this approach. The results show that the choice of the initial sample size does affect the performance of our selection approach.

Keywords—Ranking and Selection, Ordinal Optimization, Optimal Computing Budget Allocation, Subset Selection, Indifference-Zone, Initial Sample Size.

I. INTRODUCTION

WE consider optimizing the expected performance of a complex stochastic system that cannot be evaluated exactly, but has to be estimated using simulation. Our goal is to solve the following optimization problem

$$\min_{\theta \in \Theta} J(\theta) \quad (1)$$

where the feasible solution set Θ is a finite, huge and has no structure. Meanwhile, J is the expected performance measure, L is a deterministic function depends on θ and ξ , and we can write $J(\theta) = E[L(\theta, \xi)]$, θ is a vector that representing the system design parameters, and ξ represents all the random effect of the system. If we simulate the system to get estimate of $E[L(\theta, \xi)]$, then the confidence interval of this estimate cannot be improved faster than $1/\sqrt{k}$ where k is the number of samples used to get estimates of $J(\theta)$. This rate maybe good for some problems with a small number of alternatives but it is not good enough for the class of complex simulation which we consider in this paper. Thus, one could compromise the objective to get a good enough solution rather than doing extensive simulation.

Ranking and Selection ($R\&S$) procedures, are used to select the best system or a subset that contain the best systems when the number of alternatives is small, see Kim and Nelson [1]. The problem arise for a large scale problems since it needs a huge computational time. In this situation, we would compromise our objective to finding good systems rather than estimating accurately the performance value for these systems. The idea lies in Ordinal Optimization (OO) procedure, that proposed by Ho et al. [2].

In many selection procedures, sample size in the first stage t_0 play an important role to the performance of these

procedures. In fact, the initial sample size t_0 cannot be too small since we might get a poor estimates for the sample mean and variances. On the other hand, t_0 cannot be too large, because in the first stage there exist many noncritical system and by giving a large number of sample will result in losing a large number of samples and also wasting computation time. However, Chen et al. [3] and Chen et al. [4] suggested that a t_0 should be between 10 and 20 as a good choice for the initial sample size. Unfortunately, there is no clear formula to calculate an appropriate value of the initial sample size t_0 for the selection procedures, when the number of alternative is large.

In this paper, we study the effects of the initial sample size t_0 on the performance of one of the selection procedures; a sequential approach Almomani and Abdul Rahman [5]. We also consider a heuristic approach to selecting a good simulated system with high probability when the number of alternative system is huge. This approach consists of four stages; in the first stage we use the OO procedure to select randomly a subset that overlaps with the set of the actual best $m\%$ systems with high probability from the feasible solution set Θ . In the second stage, we use Optimal Computing Budget Allocation ($OCBA$) technique to allocate the available computing budget in a way that maximizes the probability of correct selection. This will follow by a Subset Selection (SS) procedure to get a smaller subset that contains the best system from the subset that is selected before. In the final stage, we use the Indifference-Zone (IZ) procedure to select the best system among the survivors in the previous stage. This approach are applied to $M/M/1$ queuing system with a different choice of the initial sample size t_0 to know the effect of the t_0 on the selection approach performance.

This paper is organized as follows; In the next section we give a background about OO , $OCBA$, SS , and IZ procedures. In Section 3, we present our sequential approach. Section 4, includes the $M/M/1$ queuing system example. Finally, in Section 5, we give some concluding remarks.

II. BACKGROUND

A. Ordinal Optimization (OO)

The OO procedure has emerged as an efficient technique for simulation and optimization. The aim of this procedure is to find good systems, rather than estimating the performance value of these systems accurately. The OO procedure has been proposed by Ho et al. [2].

Suppose that the Correct Selection (CS) is to select a subset G of g systems from the feasible solution set Θ that

contains at least one of the top $m\%$ best systems. Since we assume that Θ is very huge then the probability of CS is given by $P(CS) \approx (1 - (1 - \frac{m}{100})^g)$. Now, suppose that the CS is to select a subset G of g systems that contains at least r of the best s systems. Let S be the subset that contains the actual best s systems, then here the probability of CS can be obtained using the hypergeometric distribution as, $P(CS) = P(|G \cap S| \geq r) = \sum_{i=r}^g \frac{\binom{s}{i} \binom{n-s}{g-i}}{\binom{n}{g}}$. However, since we assumed that the number of alternatives is very large then the $P(CS)$ can be approximated by the binomial random variable. Therefore, $P(CS) \approx \sum_{i=r}^g \binom{g}{i} (\frac{m}{100})^i (1 - \frac{m}{100})^{g-i}$, where we assume that $s/n \times 100\% = m\%$. It is clear that this $P(CS)$ will increase when the sample size g increases.

B. Optimal Computing Budget Allocation (OCBA)

The *OCBA* was proposed to improve the performance of *OO* by determining the optimal numbers of simulation samples for each system, instead of equally simulating all systems. The goal of this procedure is to allocate the total simulation samples from all systems in a way that maximizes the probability of selecting the best system within a given computing budget. For more details of *OCBA* see Chen et al. [3], Chen et al. [4], and Chen [6].

Let B be the total sample that available for solving the optimization problem given in (1). Our goal is to allocate these computing simulated samples to maximize the $P(CS)$. The mathematical notation is written as

$$\begin{aligned} & \max_{T_1, \dots, T_n} P(CS) \\ & s.t. \quad \sum_{i=1}^n T_i = B \\ & \quad T_i \in \mathbf{N} \quad i = 1, 2, \dots, n \end{aligned}$$

where \mathbf{N} is the set of non-negative integers, T_i is the number of samples allocated to system i and $\sum_{i=1}^n T_i$ denotes the total computational samples and assumes that the simulation times for different systems are roughly the same. To solve this problem Chen et al. [3] proposed the following theorem.

Theorem 1: Given a total number of simulated samples B to be allocated to n competing systems whose performance is depicted by random variables with means $J(\theta_1), J(\theta_2), \dots, J(\theta_n)$, and finite variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ respectively, as $B \rightarrow \infty$, the approximate probability of CS can be asymptotically maximized when

- 1) $\frac{T_i}{T_j} = \left(\frac{\sigma_i / \delta_{b,i}}{\sigma_j / \delta_{b,j}} \right)^2$; where $i, j \in \{1, 2, \dots, n\}$ and $i \neq j \neq b$.
- 2) $T_b = \sigma_b \sqrt{\sum_{i=1, i \neq b}^n \frac{T_i^2}{\sigma_i^2}}$

where $\delta_{b,i}$ the estimated difference between the performance of the two systems ($\delta_{b,i} = \bar{J}_b - \bar{J}_i$), and $\bar{J}_b \leq \min_i \bar{J}_i$ for all i . Here $\bar{J}_i = \frac{1}{T_i} \sum_{j=1}^{T_i} \xi_{ij}$, where ξ_{ij} is a sample from ξ_i for $j = 1, \dots, T_i$.

C. Subset Selection (SS)

SS procedure screens out the search space to eliminate non-competitive systems and construct a subset that contains the best system with high probability. This procedure is suitable when the number of alternatives is relatively large, and is used to select a random subset size that contains the actual best system. It is required that $P(CS) \geq P^*$, where the Correct Selection (*CS*) is selecting a subset that contains the actual best system, and P^* is a predetermined probability.

The *SS* procedure dating back to Gupta [7], who presented a single stage procedure for producing a subset containing the best system with a specified probability. Extensions of this work which is relevant to the simulation setting include Sullivan and Wilson [8] who derived a two stage *SS* procedure that determines a subset of maximum size m that, with a specified probability will contain systems that are all within a pre-specified amount of the optimum.

D. Indifference-Zone (IZ)

The goal of *IZ* procedure is selecting the best system among n systems when the number of alternatives less than or equal 20. Suppose we have n alternative systems that are normally distributed with unknown means $\mu_1, \mu_2, \dots, \mu_n$, and suppose that these means are ordered as $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[n]}$. We want to select the system that has the best minimum mean $\mu_{[1]}$. The *IZ* is defined to be the interval $[\mu_{[1]}, \mu_{[1]} + \delta]$, where δ is a predetermined small positive real number. We are interested in selecting an alternative i^* such that $\mu_{i^*} \in [\mu_{[1]}, \mu_{[1]} + \delta]$. Let *CS* here is selecting an alternative whose mean belongs to the indifference zone. We would like the *CS* to take place with high probability, say with a probability not smaller than P^* where $1/n \leq P^* \leq 1$.

The *IZ* procedure consists of two stages. In the first stage, all systems are sampled using t_0 simulation runs to get an initial estimate of expected performance measure and their variances. Next, depending on the information obtained in the first stage, we compute how many more samples are needed in the second stage for each system to guarantee that $P(CS) \geq P^*$. Rinott [9] has presented a procedure that is applicable when the data are normally distributed and all systems are simulated independently of each others. This procedure consists of two stages for the case when variances are completely unknown. On the other hand, Tamhane [10] has presented a simple procedure that is valid when variances may not be equal.

To achieve the *CS* with high probability, *R&S* procedures needs a huge computational time, so it is not practical when n is large. Therefore the combined approaches are proposed to reduce the competent system. Nelson et al. [11] proposed a two-stage subset selection procedure. The first stage is to reduce the number of competitive systems. These systems are carried out to the second stage in which involved with the *IZ* procedure using the information gathered from the first stage. Alrefaei and Almomani [12] proposed two sequential algorithms for selecting a subset of k systems that is contained in the set of the top s systems. Another comprehensive review

of $R\&S$ procedures can be found in Bechhofer et al. [13], Goldsman and Nelson [14], and Kim and Nelson [15].

III. THE SEQUENTIAL APPROACH

Our sequential approach consists four procedures, OO , $OCBA$, SS , and IZ . Initially, using OO procedure, a subset G is randomly selected from a feasible solution set that overlaps with the set that contains the actual best $m\%$ systems with probability $(1 - \alpha_1)$. Then $OCBA$ procedure is used to allocate the available computing budget. This is follows with SS procedure to get a smaller subset I with a probability equal to $(1 - \alpha_2)$, that contains the best system among the previous selected subset. Finally, IZ procedure is applied to select the best system from that set I with probability equal to $(1 - \alpha_3)$.

Algorithm:-

Setup: Specify g where $|G| = g$, k where $|G'| = k$, the number of initial simulation samples $t_0 \geq 2$, the indifference zone δ , and $t = t_{(1-\alpha_2/2)^{\frac{1}{g-1}}, t_0-1}$ from the t -distribution. Let $T_1^l = T_2^l = \dots = T_g^l = t_0$, and determine the total computing budget B . Here, G is the selected subset from Θ , that satisfies $P(G \text{ contains at least one of the best } m\% \text{ systems}) \geq 1 - \alpha_1$, whereas G' is the selected subset from G , where $g \geq k$. The iteration number is represented by l .

Select a subset G of size g randomly from Θ . Take a random samples of t_0 observations y_{ij} ($j = 1, \dots, t_0$) for each system i in G , where $i = 1, \dots, g$.

Initialization: Calculate the sample mean and variances

$$\bar{y}_i^{(1)} \text{ and } s_i^2, \text{ where } \bar{y}_i^{(1)} = \frac{\sum_{j=1}^{T_i^l} y_{ij}}{T_i^l} \text{ and } s_i^2 = \frac{\sum_{j=1}^{T_i^l} (y_{ij} - \bar{y}_i^{(1)})^2}{T_i^l - 1}, \text{ for all } i = 1, \dots, g.$$

Order the systems in G according to their sample averages; $\bar{y}_{[1]}^{(1)} \leq \bar{y}_{[2]}^{(1)} \leq \dots \leq \bar{y}_{[g]}^{(1)}$. Then select the best k systems from the set G , and represent this subset as G' .

Stopping Rule: If $\sum_{i=1}^g T_i^l \geq B$, then stop. Otherwise, randomly select a subset G'' of the $g - k$ alternatives from $\Theta - G'$, let $(G = G' \cup G'')$.

Simulation Budget Allocation: Increase the computing budget by Δ and compute the new budget allocation, $T_1^{l+1}, T_2^{l+1}, \dots, T_g^{l+1}$, by using Theorem 1.

Perform additional $\max\{0, T_i^{l+1} - T_i^l\}$ simulations for each system i , $i = 1, \dots, g$, let $l \leftarrow l + 1$. Go to **Initialization**.

Screening: Set $I = \{i : 1 \leq i \leq k \text{ and } \bar{y}_i^{(1)} \geq \bar{y}_j^{(1)} - [W_{ij} - \delta]^-, \forall i \neq j\}$, where $W_{ij} = t \left(\frac{s_i^2}{T_i} + \frac{s_j^2}{T_j} \right)^{1/2}$ for all $i \neq j$, and $[x]^- = x$ if $x < 0$ and $[x]^- = 0$ otherwise.

If I contains a single index, then this system is the best system. Otherwise, for all $i \in I$, compute the second sample size $N_i = \max\{T_i, \lceil (\frac{hs_i}{\delta})^2 \rceil\}$, where $h = h(1 - \alpha_3/2, t_0, |I|)$ be the Rinott [9] constant and can be obtained from tables of Wilcox [16].

Take additional $N_i - T_i$ random samples of y_{ij} for each system $i \in I$, and compute the overall sample

$$\text{means for } i \in I \text{ as } \bar{y}_i^{(2)} = \frac{\sum_{j=1}^{N_i} y_{ij}}{N_i}$$

Select system $i \in I$ with the smallest $\bar{y}_i^{(2)}$ as the best.

Remarks:-

- The initial sample size t_0 is the number of observations that are taken in the first stage (OO procedure) in order to get an initial estimate of mean and variance for each system. Note that, if t_0 is too small, we might get a poor estimate of σ_i^2 (s_i^2). In particular, it could be that s_i^2 is much greater than σ_i^2 , leading to an unnecessarily large value of N_i .
- The Rinott constant $h = h(1 - \alpha_3/2, t_0, |I|)$ is determined by the desired confidence level $(1 - \alpha_3/2)$, the initial sample size t_0 , and the number of systems in the set I ($|I|$). From tables of Wilcox [16] we note that, the constant h increases in $|I|$, and decreases in α_3 and t_0 . The experiment design factor that is under control is t_0 .
- Nelson et al. [11] have shown that with probability at least $1 - (\alpha_2 + \alpha_3)$ our sequential approach selects the best system in the subset G . Therefore, if G contains at least one of the top $m\%$ systems, then our approach selects a good system with probability $1 - (\alpha_2 + \alpha_3)$. On the other hand, from the OO procedure we can show that the selected set G , contains at least one of the best $m\%$ systems with probability $(1 - \alpha_1) = 1 - (1 - \frac{m}{100})^g$. Therefore, $P(\text{the selected system in the sequential approach is in the top } m\% \text{ systems}) \geq (1 - (1 - \frac{m}{100})^g)(1 - (\alpha_2 + \alpha_3)) \geq 1 - ((1 - \frac{m}{100})^g + \alpha_2 + \alpha_3)$.

IV. NUMERICAL EXAMPLE

In our example, we consider the $M/M/1$ queuing systems where the inter arrival times and the service times are exponentially distributed and the system has one server. Our goal is selecting one of the best $m\%$ systems that has the minimum average waiting time per customer from n $M/M/1$ queuing systems.

Moreover, as a measure of selection quality, we use the Probability of Correct Selection ($P(CS)$), and the Expected Opportunity Cost ($E(OC)$) of a potentially incorrect selection, where the Opportunity Cost (OC) is the difference between unknown mean of the selected best system and the actual best system. More details of $E(OC)$ can be found in He et al. [17], and Chick and Wu [18]. In our approach, we consider the $E(OC)$ as the absolute value of the difference of unknown mean between the selected best system and the actual best system.

We apply our sequential approach in this example under some assumptions. We assume that the arrival rate λ is constant and the service rate μ is belong to the interval $[a, b]$ and particularly, take $\lambda = 1$ and $\mu \in [4, 5]$. Suppose that we have 1000 of $M/M/1$ queuing systems, and we discretize the problem by assuming that $\Theta = \{4.001, 4.002, \dots, 5.000\}$. Therefore, the best queuing system would be the 1000th queuing system with $\mu_{1000} = 5.0$. Let $n = 1000$, $g = 50$, $\alpha_2 = \alpha_3 = 0.005$, $\delta = 0.05$, $k = 10$ and $\Delta = 20$ (these settings are chosen arbitrarily). Suppose we want to select one of the best (5%) systems, then our target is the systems from

951 to 1000. Here the correct selection would be selecting the system that belongs to $\{\mu_{951}, \mu_{952}, \dots, \mu_{1000}\}$, and the analytical Probability of the Correct Selection can be calculated as $P(CS) \geq 1 - \left(\left(1 - \frac{5}{100}\right)^{50} + 0.005 + 0.005 \right) \geq 0.91$.

Meanwhile, to study the effect of the initial sample size t_0 on our approach, we choose five different values for the t_0 as 10, 20, 30, 50 and 100. In the first experiment, we consider the total number of simulation samples; "Total Budget" $B = 1300$ for all values of t_0 . We find that our approach works just fine when $t_0 = 10, 20$ and it is fail to working when $t_0 = 30, 50, 100$. Following this in the next experiment, we consider the minimum value of B for all cases of t_0 , to see the effects of t_0 on performance of our approach.

Table I contains the result of this experiment, over 100 replications for selecting one of the best (5%) systems. From the table, " $\min B$ " is the minimum value of the total budget, $\sum_{i=1}^g T_i$ is the average number of the total sample size in **Stopping Rule** in our algorithm, and $\sum_{i \in I} N_i$ is the average number of the total sample size in **Screening** step in our algorithm, and $E(OC)$ is the average number of Expected Opportunity Cost.

From Table I, we note that, the first parameter that affected by the initial sample size t_0 is the total budget B . For at least the minimum value of B for each t_0 our algorithm work and move from the first stage to the next stage. Moreover, we note that $\sum_{i=1}^g T_i$ are keep changing for different value of t_0 with increasing pattern likewise t_0 . This is expected since T_i is related to B . Furthermore, $\sum_{i \in I} N_i$ approximately the same in all cases of t_0 except when $t_0 = 10$. This happened because in this approach we calculate the values of N_i after we increase the Δ and compute the new budget allocation for each system. Clearly, $\sum_{i \in I} N_i$ is not related to B . Actually, the effect of t_0 on N_i is almost non exist when t_0 relatively large, but in the other hand for a small value of t_0 , it will end up with larger value of N_i .

Besides that, from Table I we can see that the $P(CS)$ for our approach is closed to the analytical $P(CS)$. It shows that the high $P(CS)$ occurs when $t_0 = 10$ with the value as high as 84% comparing with the other values of t_0 . However, the important note here is that when t_0 is 10 we get the lowest value for the $E(OC)$. Also that, for all values of t_0 the $E(OC)$ between our approach and the analytical $E(OC)$ are closed together, except when $t_0 = 10$. For this initial value ($t_0 = 10$), there is a huge difference in the value of $E(OC)$ between our approach and the analytical values. This happened because when t_0 is small, we tend to get a poor estimates for the first mean and variance.

V. CONCLUSION

In this paper we discuss the effect of the initial sample size t_0 on the performance of a sequential approach that is used to selecting a good simulated system, when the number of alternatives is large. This approach consists four stage. Initially, using *OO* procedure, a subset G is randomly selected form a feasible solution set that overlap with the set that contains the actual best $m\%$ systems with high probability. Then *OCBA* procedure is used to allocate the available

computing budget. This is follows with *SS* procedure to get a smaller subset I with high probability, that contains the best system among the previous selected subset, where $|I| \leq 20$. Finally, *IZ* procedure is applied to select the best system from that set I . We apply this approach in *M/M/1* queuing system under some parameters setting, and we test five different value for t_0 , to study the effect of t_0 on our approach. We note that our approach is affected by t_0 in different rooms. The main parameters affected by t_0 are, the total budget B , the total sample size in **Stopping Rule** T_i , and the Expected Opportunity Cost ($E(OC)$) of a potentially incorrect selection. From the numerical result we note that the initial sample size t_0 here affect on B , where we need the minimum value of B to make the approach work. Also we note that the value of T_i increases when t_0 increase. Finally, for the small value of t_0 will end up with high value in ($E(OC)$). Since the performance of our sequential approach is sensitive to the initial sample size t_0 , in a future work a "zeroth stage" of sampling should be added to the approach in order to determine an adequate choice for the value of the initial sample size t_0 for each system.

ACKNOWLEDGMENT

The authors would like to thank School of Mathematical Sciences, Universiti Sains Malaysia for supporting this research.

REFERENCES

- [1] S.H. Kim and B.L. Nelson, "Selecting the best system," *Operations Research and Management Science*, Chapter 17, pp. 501–534, 2006.
- [2] Y.C. Ho, R.S. Sreenivas and P. Vakili, "Ordinal optimization of DEDS," *Journal of Discrete Event Dynamic Systems*, vol. 2, pp. 61–88, 1992.
- [3] C.H. Chen, E. Yücesan and S.E. Chick, "Simulation budget allocation for further enhancing the efficiency of ordinal optimization," *Discrete Event Dynamic Systems*, vol. 10, no. 3, pp. 251–270, 2000.
- [4] C.H. Chen, C.D. Wu and L. Dai, "Ordinal comparison of heuristic algorithms using stochastic optimization," *IEEE Transaction on Robotics and Automation*, vol. 15, no. 1, pp. 44–56, 1999.
- [5] M.H. Almomani and R. Abdul Rahman, "Selecting a good stochastic system for the large number of alternatives," Accepted and will be published in *Communications in Statistics: Simulation and Computation*, 2011.
- [6] C.H. Chen, "An effective approach to smartly allocated computing budget for discrete event simulation," *IEEE Conference on Decision and control*, vol. 34, pp. 2598–2605, 1995.
- [7] S.S. Gupta, "On some multiple decision (selection and ranking) rules," *Technometrics*, vol. 7, no. 2, pp. 225–245, 1965.
- [8] D.W. Sullivan and J.R. Wilson, "Restricted subset selection procedures for simulation," *Operations Research*, vol. 37, pp. 52–71, 1989.
- [9] Y. Rinott, "On two-stage selection procedures and related probability-inequalities," *Communications in Statistics: Theory and Methods*, vol. A7, pp. 799–811, 1978.
- [10] A.C. Tamhane and R.E. Bechhofer, "A two-stage minimax procedure with screening for selecting the largest normal mean," *Communications in Statistics: Theory and Methods*, vol. A6, pp. 1003–1033, 1977.
- [11] B.L. Nelson, J. Swann, D. Goldsman and W. Song, "Simple procedures for selecting the best simulated system when the number of alternatives is large," *Operations Research*, vol. 49, pp. 950–963, 2001.
- [12] M.H. Alrefaei and M.H. Almomani, "Subset selection of best simulated systems," *Journal of the Franklin Institute*, vol. 344, no. 5, pp. 495–506, 2007.
- [13] R.E. Bechhofer, T.J. Santner and D.M. Goldsman, *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: Wiley, 1995.
- [14] D. Goldsman and B.L. Nelson, "Ranking, selection and multiple comparisons in computer simulation," *Proceedings of the 1994 Winter Simulation Conference*, pp. 192–199.

- [15] S.H. Kim and B.L. Nelson, "Recent advances in ranking and selection," *Proceedings of the 2007 Winter Simulation Conference*, pp. 162–172.
- [16] R.R. Wilcox, "A table for Rinott's selection procedure," *Journal of Quality Technology*, vol. 16, pp. 97–100, 1984.
- [17] D. He, S.E. Chick and C.H. Chen, "Opportunity cost and OCBA selection procedures in ordinal optimization for a fixed number of alternative systems," *IEEE Transactions on Systems*, vol. 37, pp. 951–961, 2007.
- [18] S.E. Chick and Y. Wu, "Selection procedures with frequentist expected opportunity cost bounds," *Operations Research*, vol. 53 no. 5, pp. 867–878, 2005.

TABLE I
THE NUMERICAL ILLUSTRATION FOR
 $n = 1000, g = 50, m\% = 5\%, k = 10, \Delta = 20$

t_0	$\min B$	$\sum_{i=1}^g T_i$	$\sum_{i \in I} N_i$	$P(CS)$		$E(OC)$	
				Suggested approach	Analytical	Suggested approach	Analytical
10	500	1880	3920	84%	91%	0.012408734	0.004870680
20	1000	2805	2767	78%	91%	0.008835964	0.008004247
30	1500	3765	2680	78%	91%	0.009004282	0.008618002
50	2500	5600	2553	76%	91%	0.006561811	0.009390408
100	5000	10282	2750	75%	91%	0.005444701	0.008958827