STUDY OF PROPERTIES OF STATIONARY PROBABILITY DISTRIBUTIONS ARISING IN IMPLEMENTING SELF ORGANIZING SCHEMES TO LINEAR STRUCTURE

DR. PRATIKSHA BHAWALKAR, DR. MEENAKSHI M. SAGDEO

Abstract: In this computers era one hasto deal with enormous electronic data. For retrieval of desired data differentsearch techniques are used, very basic of them is linear search technique. While using LST (Linear Search Technique) the search cost is directly proportional to the position of record in the list. To overcome this drawback various Self Organizing Schemes (SOS) were devised by different scientists over the last 5 decades. Self-organizingscheme is a technique of rearranging records after each search, so that in the long run more frequently requested records will be placed in the beginning of the list. In this paper, we consider a special request probability distribution for n records, where only one record is the most important with request probability p and the remaining (n-1) records have equal request probability, say, r. (p+(n-1)r=1 and p > r > 0). We find the stationary probability distribution for the position of the most important record in the long run. This gives us idea about the expected search cost. We then compare various Self-organizing schemes like Transposition, Move-to-front, Modified-Move-to-Front, POSand SWITCHon the basis of properties of stationary probability distributions.

Keywords: Linear Search, Move-to-Front Scheme, Optimal Ordering, Self-organizing Schemes

Introduction: In this paper we study the properties of stationary probability distributions arising after implementing different SOS. We consider all SOS under a special request probability distribution, where only one of the 'n' records is the most important, having relatively very high request probability 'p' and the remaining (n-1) records have the same request probability 'r'; such that p+(n-1)r= 1. The stationary probability distributions are obtained in [1] and [2]. In section 1.1 we briefly define various SOS, in section 1.2 we define random variable X, which is the position of the most important record in the long run, in section 1.3 we give the stationary probability distributions of X for various SOS. In section 1.4, we present procedure for generating new sample, in section 1.5 we give notations used for the R commands, in section 1.6 we discuss use of R commands to find measures of central tendency and dispersion for different SOS. We also explain the R-commands used to plot p.m.f., c.d.f. and histogram. In section 1.7, we present observationsand finally in section 1.8 we draw conclusions.

Different Self Organizing Schemes:

1. **MTF** is a self-organizing scheme where the requested record is placed at the first position after use and all other records are shifted/ moved towards end of the list to make room for newly requested record. If the requested record is at the first position it retains its position.
2. **TR** is a self-organizing scheme where the requested record after used interchanges its position with the immediately preceding record. If the requested record is at the first position it retains its position.
3. **POS(k)**: If the record is requested from the first position, then it will retain its position.If the record is requested from the position 'j' (j= 2 to k) it will be interchanged with the immediately preceding record and all other records will remain unaltered.If the record is requested from the position 'j' (j= k+1, ..., n) then it will be placed at k^{th} position by shifting the records originally in the positions k to (j - k) one position backward to make room for it.
4. **SWITCH (k)**:If the record is requested from the first position, then it will retain its position.If the record is requested from the position 'j' (j= 2 to k) it will be placed at position '1'by shifting the records originally in the positions 1 to (j - 1) one position backward to make room for it. If the record is requested from the position 'j' (j= k+1, ..., n) then it is interchanged with immediate preceding record (j – 1) and all other records remain unaltered.
5. **MMTF (k)**: If the record is requested from the first position, then it will retain its position. If the record is requested from the position 'j' (j = 2, ..., k) it will be placed at position '1' by shifting the records originally in the positions 1 to (j - 1), one position backward to make room for it. If the record is requested from the position 'j' (j = k+1, ..., n) it will be placed at position 'k' by shifting the records originally in the positions k to (j - 1), one position backward to make room for it.

Definition of Random Variable under Consideration: We define random variable 'X' as the position of the most important record in the long run.

Stationary Probability Distributions of X for Various SOS:

For TR Scheme

$$\pi_1 = \frac{\left(1-\frac{r}{p}\right)}{\left(1-(\frac{r}{p})^n\right)}$$

$$\pi_j = \left(\frac{r}{p}\right)^{j-1} \pi_1 \quad \text{for } j = 2 \text{ to } n$$

For MTF Scheme

$\pi_j = p$ for j=1

$$\pi_j = \frac{p(1-p)(1-p-r)...(1-p-(j-2)r)}{(1-r)(1-2r)...(1-(j-1)r)} \quad \text{for } j = 2 \text{ to } n$$

For MMTF Scheme

(i) $\pi_j = A_j\pi_{j-1}$

where $A_j = \frac{[(k+1)-j].r}{p+(k-j).r}$ for $j = 2, \dots, k$

(ii) $\pi_j = B_s\pi_{s-1}$

where $B_s = \frac{[(n+1)-s].r}{p+(n-s).r}$ for $s = (k+1), \dots, n$

(iii) π_1 is obtained by $\pi_1 = 1 - \sum_{j=2}^{n} \pi_j$

These stationary probability distributions are obtained in [1] and [2].

Generation of Random Sample: To study the properties of stationary probability distribution of 'X' we generate random a sample from it.

In C++ program, we use random() function for generating a random number.

If $[\pi_1, \pi_2, \dots \dots \dots, \pi_n]$ is the stationary probability distribution, and the random number is less than or equal toπ_1, random variable X will take value 1. Similarly, if random number generated is greater than π_1 and is less than or equal to$\pi_1 + \pi_2$, then X will take value 2. In this way, we generate random sample. In the following section, we give the notations used in writing R- Commands.

Notation:

X : Position of the most important record in the long run.

i : Value taken by X where i = 1, ..., n

f_i : Frequency corresponding to X = i

p(x) : Probability mass function of X=P[X=x] for x= 1, 2, ..., n

E(X) : Expected search cost in the long run for the most important record $= \sum_{i=1}^{n} i\,\pi_i$, where π_i is the stationary probability that the most important record is in position i.

m : Sample size

$F_X(x)$: Probability distribution function of X.

R–Commands: We considera random sample of size 'm' from a stationary probability distribution of X for a particular SOS. On the basis of the frequencies for different values of X, we find probability mass function(pmf) of X. We then find cumulative distribution function of X using pmf. By writing R-Commands summary( ) and var ( ), we find measures of central tendency and measures of dispersions. We use functions pmf( ), cdf( ) and histogram( ) in R to plot probability mass function, cumulative function and histogram.

Observations: From the **Table 1.1**, we observe that for fixed values of 'n', 'p' and 'k',

$[\text{Mean}]_{MTF} \geq [\text{Mean}]_{MMTF} \geq [\text{Mean}]_{POS} \geq [\text{Mean}]_{SWITCH} \geq [\text{Mean}]_{TR}$

From the **Table 1.2**, we observe that probability distribution functions for TR, MMTF, MTF, POS and SWITCH scheme for fixed values of 'n', 'p' and 'k', satisfy the following inequality

$[P(X\leq x)]_{MTF} \leq [P(X\leq x)]_{MMTF} \leq [P(X\leq x)]_{POS} \leq [P(X\leq x)]_{SWITCH} \leq [P(X\leq x)]_{TR}$

Thus, for fixed 'n', 'p' and 'k' we get the following result

$[\text{Median}]_{MTF} > [\text{Median}]_{MMTF} > [\text{Median}]_{POS} > [\text{Median}]_{SWITCH} > [\text{Median}]_{TR}$

Table 1.1: Measures of Central Tendencyand Measure of Dispersion of X (n = 10; p = 0.2; k=3; m = 500)

	TR	POS	SWITCH	MMTF	MTF
Mean	1.506	2.244	1.612	2.292	2.818
Mode	1	1	1	1	1
Median	1	1	1	1	2
Q1	1	1	1	1	1
Q3	2	3	2	3	4
Range	6	9	8	9	9
Variance	0.87972	4.07088	5.17122	1.33212	4.10868

Table 1.2: Probability Distribution Function of Xfor Different SOS(n=10; p=0.2; k=3; m=500)

X TR	1	2	3	4	5	6	7	≥ 8
fi	355	72	48	18	5	1	1	0
p(x)	0.71	0.144	0.096	0.036	0.01	0.002	0.002	0
$F_X(x)$	0.71	0.854	0.95	0.986	0.996	0.998	1	1

X MMTF	1	2	3	4	5	6	7	8	9	10
fi	287	71	37	29	20	27	13	9	5	2
p(x)	0.574	0.142	0.074	0.058	0.04	0.054	0.026	0.018	0.01	0.004
$F_X(x)$	0.574	0.716	0.79	0.848	0.888	0.942	0.968	0.986	0.996	1

XPOS	1	2	3	4	5	6	7	8	9	10
fi	300	65	39	16	25	23	17	7	7	1
p(x)	0.6	0.13	0.078	0.032	0.05	0.046	0.034	0.014	0.014	0.002
$F_X(x)$	0.6	0.73	0.808	0.84	0.89	0.936	0.97	0.984	0.998	1

XSWITCH	1	2	3	4	5	6	7	8	9	10
fi	337	90	37	18	8	7	1	1	1	0
p(x)	0.674	0.18	0.074	0.036	0.016	0.014	0.002	0.002	0.002	0
$F_X(x)$	0.674	0.854	0.928	0.964	0.98	0.994	0.996	0.998	1	1

XMTF	1	2	3	4	5	6	7	8	9	10
fi	228	64	55	43	32	34	19	9	13	3
p(x)	0.456	0.128	0.11	0.086	0.064	0.068	0.038	0.018	0.026	0.006
$F_X(x)$	0.456	0.584	0.694	0.78	0.844	0.912	0.95	0.968	0.994	1

Figure 1.1
Probability Mass Function of X
For TR, MTF, MMTF, POS and SWITCH
(n=10; p=0.2; k=3; m=500)

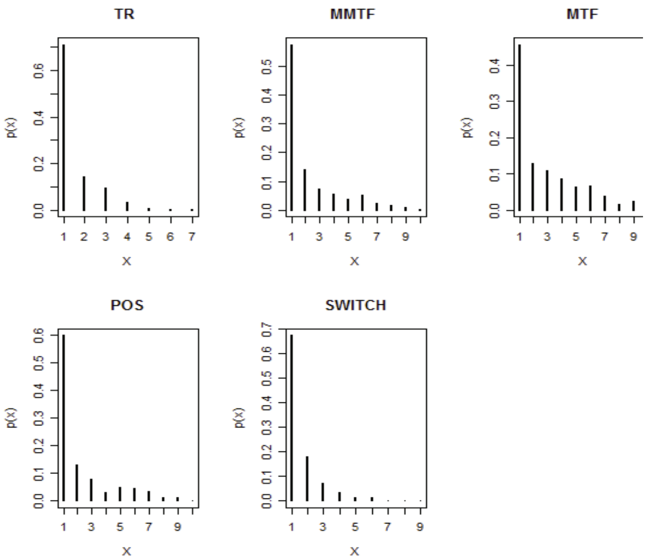


Figure 1.2
Cumulative Distribution function of X
For TR, MTF, MMTF, POS and SWITCH
(n=10;p=0.2;k=3;m=500)

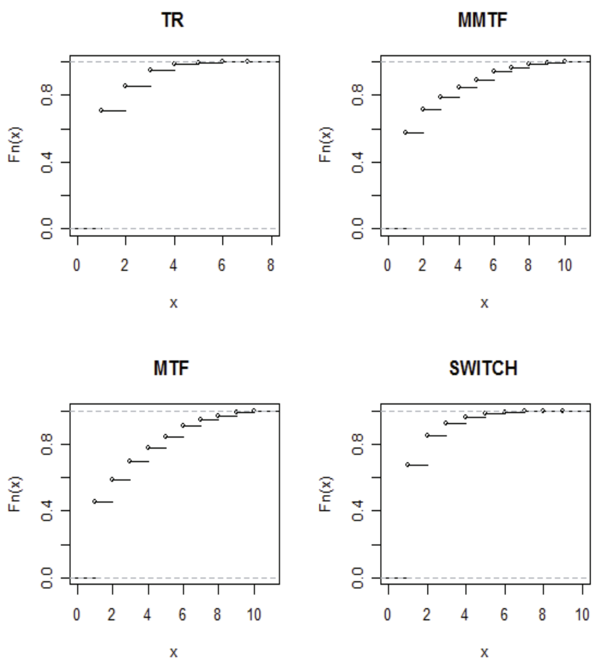
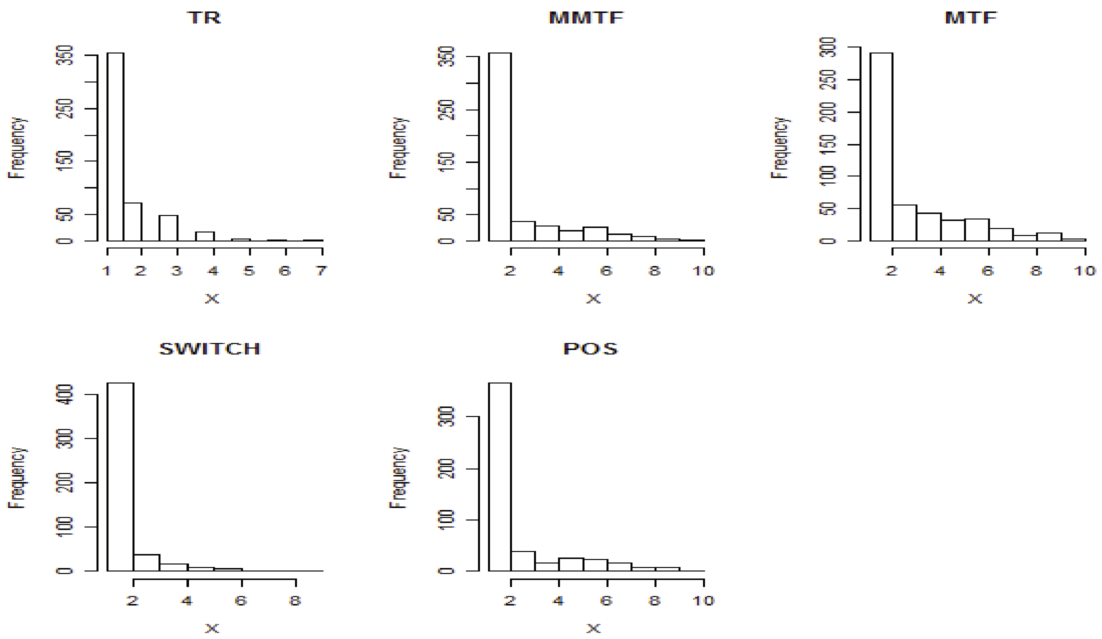


Figure 1.3
Histogram of X For TR, MTF, MMTF, POS and SWITCH (n=10;p=0.2;k=3;m=500)



Conclusions: From the above results, we can say that for fixed number of records 'n', request probability for the most important record 'p' and the splitting point 'k', the median for MTF scheme is larger than that for TR scheme. Hence,ASC for MTF Scheme is higher than that of TR scheme.

$$ASC_{TR} < ASC_{SWITCH} < ASC_{POS} < ASC_{MMTF} < ASC_{MTF} \quad ...(1)$$

In the literature, various SOS are compared using expected average search cost and rate of convergence. It was shown by Rivest[4] that average search cost of TR is less than that of MTF. We have obtained the same result while comparing different SOS using random variable 'X',the position of most important record in the long run.This result is also proved inBhawalkar and Sagdeo [1]. MMTF(k) is a hybrid scheme and italways performs better than MTF. Thus, use of hybrid scheme MMTF is preferable over the MTF scheme to reduce average search cost.

References:

1. *K. Hemlatha, M.V.D.N.S. Madhavi, D. Rajani,* Effect Of Melting On Non-Darcy MHD Mixed Convective Flow From A Vertical Plate Embedded In A Saturated Porous Medium. ; Mathematical Sciences International Research Journal : ISSN 2278-8697 Volume 5 Spl Issue (2016), Pg 74
2. Bhawalkar and Sagdeo (2012), A Modified Move to Front Scheme, Vidarbha Journal of science, Vol. 7, No. 1-2, p-20-32.
3. Bhawalkar and Sagdeo (2012), A New Criterion for Comparison of Self Organizing Schemes, Mathematical Sciences International Research Journal, Vol 1, No. 1, p-303-311.
4. McCabe J. (1965), On Serial Files with Relocatable Records Operations Research 12, 609-618.
5. Rivest R. (1976) On Self Organizing Sequential Search Heuristics. Comm. Of the A C.M., p 63 -67
6. Tenenbaum, A. (1978). Simulations of Dynamic Sequential Search Algorithms. Commun. ACM 21, 9 (Sept.), 790-791.
7. *N. Swathi, T. Sumathi Uma Maheswari, M. Tirumala Devi,* Reliability Analysis Of Time – Dependent System When The Number Of Cycles Follow Binomial Distribution And Stress- Strength Follow Pareto Distribution; Mathematical Sciences International Research Journal : ISSN 2278-8697 Volume 5 Spl Issue (2016), Pg 45-58
8. Tenenbaum, A. And Nemes, R. M. (1982) Two Spectra of Self Organizing Sequence Search Algorithm. SIAM J. COMPUT 11, 3(Aug)., 557-566

Dr. Mrs. Pratiksha Bhawalkar
Assistant Professor, Thakur Institute of Management Studies & Research, Kandivali (East), Mumbai-101.
Dr. Mrs. Meenakshi M. Sagdeo
Associate Professor, Elphinstone College Fort, Mumbai-400032.