Probability in the Engineering and Informational Sciences, **13**, 1999, 377–385. Printed in the U.S.A.

ON A CONJECTURE OF FILL AND HOLST INVOLVING THE MOVE-TO-FRONT RULE AND CACHE FAULTS

Martin Hildebrand
Department of Mathematics and Statistics
State University of New York
University at Albany
Albany, NY 12222
martinhi@math.albany.edu

Fill and Holst conjectured for the move-to-front rule that the probability that the search time is greater than c will be Schur concave in the stationary distribution for any value of c. This paper disproves the conjecture but proves some conclusions that would be implied by the conjecture.

1. INTRODUCTION

In Fill and Holst [2], the authors considered the move-to-front rule for self-organizing lists. In this list, there are n objects, and object number i has a selection probability $p_i > 0$ such that $\sum_{i=1}^{n} p_i = 1$. In the discrete version, an object is selected from the list every unit of time; in the continuous version, an object is selected from the list at times determined by a Poisson process with parameter 1. In either case, the object is selected according to its selection probability and independently of earlier selections, and the selected object is moved to the top of the list. The discrete version gives a Markov chain on the orderings of the list (which can be viewed as the symmetric group S_n) while the continuous version gives a Markov process on these orderings. In either case, the stationary distribution is given by

$$P(\sigma(\infty) = \sigma) = p_{\sigma_1} \frac{p_{\sigma_2}}{(1 - p_{\sigma_1})} \cdots \frac{p_{\sigma_n}}{(1 - p_{\sigma_1} - \cdots - p_{\sigma_{n-1}})}$$

 0269-9648/99 $12.50

https://doi.org/10.1017/S0269964899133084 Published online by Cambridge University Press

where $\sigma(\infty)$ is a random variable on S_n distributed according to this stationary distribution. (See Fill and Holst [2] or Hendricks [3].) We define the search cost to be the number of objects above the selected object.

Let $S(\infty)$ be a random variable representing the search cost of the move-to-front rule where the initial distribution of the objects is determined according to the stationary distribution of the Markov chain or process. As in [2], the convention is that $S(\infty)$ is 0 if the selected object is on top of the list. Considering computer science applications, Fill and Holst [2] note a connection between the event $S(\infty) \geq c$ and a cache fault if the size of the cache is c.

$E(S(\infty))$ is a symmetric function of $p_1, \ldots, p_n$, and Fill and Holst [2] prove the Schur concavity of $E(S(\infty))$. Marshall and Olkin [4] have an extensive discussion of Schur convexity and concavity. For our purposes, it will suffice to note that a symmetric function ϕ defined on the subset of $\mathbf{R}^n$ such that all coordinates are positive and sum to 1 will be Schur convex (or Schur concave) if $\phi(x_1, s - x_1, x_3, \ldots, x_n)$ is nondecreasing (or nonincreasing) in x_1 for $x_1 \geq s/2$ for each fixed $s, x_3, \ldots, x_n$. $P(S(\infty) \geq c)$ is also a symmetric function of $p_1, \ldots, p_n$ for each value c, and, in a remark, Fill and Holst [2] conjecture that $P(S(\infty) \geq c)$ is also Schur concave.

For some values of c, the conjecture holds. In particular, we can show

THEOREM 1: *$P(S(\infty) \geq c)$ is Schur concave if $c \in \{1,2,3\}$.*

However, we will show

THEOREM 2: *$P(S(\infty) \geq c)$ is not Schur concave if $c = 4$ and $n = 5$.*

Theorem 2 disproves the conjecture of Fill and Holst. The following theorem, which would be a corollary of the conjecture, still holds.

THEOREM 3: *For any c and any n, the maximum of $P(S(\infty) \geq c)$ occurs when $p_i = 1/n$ for $i = 1, \ldots, n$.*

2. CONJECTURE FOR $c = 1$ AND $c = 2$

In this section, we shall prove Theorem 1 in the case $c = 1$ and $c = 2$. The case $c = 3$ is tedious and omitted. The cases $c = 1$ and $c = 2$ are relatively straightforward, and Fill [1] believes he showed these cases in unpublished work exploring the conjecture.

PROOF OF THEOREM 1: The case $c = 1$ is equivalent to showing that $P(S(\infty) = 0)$ is Schur convex. We assume $p_1 \geq p_2 \geq \cdots \geq p_n$. Note that $P(S(\infty) = 0) = \sum_{i=1}^n p_i^2$ since the probability that the ith object is on top in the stationary distribution is p_i.

Suppose $i > j$ and $0 < \Delta < p_j$. Consider changing the probabilities of selecting objects i and j to $p_i + \Delta$ and $p_j - \Delta$, respectively. Leave all other selection probabilities unchanged. Since

$$\begin{aligned}(p_i + \Delta)^2 + (p_j - \Delta)^2 &= p_i^2 + p_j^2 + 2\Delta(p_i - p_j) + 2\Delta^2 \\ &> p_i^2 + p_j^2,\end{aligned}$$

we get the Schur convexity of $P(S(\infty) = 0)$.

Showing that $P(S(\infty) \geq 2)$ is Schur concave is equivalent to showing that $P(S(\infty) \leq 1)$ is Schur convex. Again, we assume that $p_1 \geq p_2 \geq \cdots \geq p_n$.

Observe that

$$\begin{aligned} P(S(\infty) \leq 1) &= P(S(\infty) = 0) + P(S(\infty) = 1) \\ &= \sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} p_j \frac{p_i}{1 - p_j} p_i \\ &= \sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} p_i^2 \frac{p_j}{1 - p_j}. \end{aligned}$$

Note that the probability (in the stationary distribution) of having the top two objects being objects j and i (in that order) is $p_j(p_i/(1 - p_j))$, and the expression for $P(S(\infty) = 1)$ follows.

Suppose $1 \leq i_1 < i_2 \leq n$ and we change the probabilities of selecting objects i_1 and i_2 to $p_{i_1} + \Delta$ and $p_{i_2} - \Delta$, respectively, where $0 < \Delta < p_{i_2}$ while leaving other selection probabilities unchanged.

The case where $j \notin \{i_1, i_2\}$ and $i \in \{i_1, i_2\}$ in the expression for $P(S(\infty) = 1)$ is handled as in the case where $c = 1$. In particular,

$$((p_{i_1} + \Delta)^2 + (p_{i_2} - \Delta)^2) \frac{p_j}{1 - p_j} \geq (p_{i_1}^2 + p_{i_2}^2) \frac{p_j}{1 - p_j}.$$

Next let us consider the case where $i \notin \{i_1, i_2\}$ and $j \in \{i_1, i_2\}$ in the expression for $P(S(\infty) = 1)$. Observe that

$$\begin{aligned} &p_i^2 \left(\frac{p_{i_1} + \Delta}{1 - p_{i_1} - \Delta} + \frac{p_{i_2} - \Delta}{1 - p_{i_2} + \Delta} \right) \\ &\quad = p_i^2 \left(-1 + \frac{1}{1 - p_{i_1} - \Delta} - 1 + \frac{1}{1 - p_{i_2} + \Delta} \right). \end{aligned}$$

Note that

$$\begin{aligned} &\frac{d}{d\Delta} \left(\frac{1}{1 - p_{i_1} - \Delta} + \frac{1}{1 - p_{i_2} + \Delta} \right) \\ &\quad = \frac{1}{(1 - p_{i_1} - \Delta)^2} - \frac{1}{(1 - p_{i_2} + \Delta)^2} \\ &\quad \geq 0 \end{aligned}$$

if $0 \leq \Delta < p_{i_2}$ and $p_{i_1} \geq p_{i_2}$. Thus for $\Delta \in [0, p_{i_2})$, the expression

$$\frac{1}{1 - p_{i_1} - \Delta} + \frac{1}{1 - p_{i_2} + \Delta}$$

https://doi.org/10.1017/S0269964899133084 Published online by Cambridge University Press

is minimized at $\Delta = 0$. Thus

$$p_i^2\left(\frac{p_{i_1}+\Delta}{1-p_{i_1}-\Delta}+\frac{p_{i_2}-\Delta}{1-p_{i_2}+\Delta}\right) \geq p_i^2\left(\frac{p_{i_1}}{1-p_{i_1}}+\frac{p_{i_2}}{1-p_{i_2}}\right).$$

If $i, j \notin \{i_1, i_2\}$, then $p_i^2(p_j/(1-p_j))$ is unchanged by the change of selection probabilities for objects i_1 and i_2.

Now suppose $i, j \in \{i_1, i_2\}$. Here we shall consider terms coming from the expression for $P(S(\infty)=0)$ as well as the expression for $P(S(\infty)=1)$. Let $m = 1 - p_{i_1} - p_{i_2}$. Observe

$$\begin{aligned}
&(p_{i_1}+\Delta)^2\left(1+\frac{p_{i_2}-\Delta}{1-p_{i_2}+\Delta}\right)+(p_{i_2}-\Delta)^2\left(1+\frac{p_{i_1}+\Delta}{1-p_{i_1}-\Delta}\right)\\
&\quad = (p_{i_1}+\Delta)^2\,\frac{1}{1-p_{i_2}+\Delta}+(p_{i_2}-\Delta)^2\,\frac{1}{1-p_{i_1}-\Delta}\\
&\quad = \frac{p_{i_1}+\Delta}{p_{i_1}+m+\Delta}\,(p_{i_1}+\Delta)+\frac{p_{i_2}-\Delta}{p_{i_2}+m-\Delta}\,(p_{i_2}-\Delta)\\
&\quad = \left(1-\frac{m}{p_{i_1}+m+\Delta}\right)(p_{i_1}+\Delta)+\left(1-\frac{m}{p_{i_2}+m-\Delta}\right)(p_{i_2}-\Delta)\\
&\quad = p_{i_1}+\Delta+p_{i_2}-\Delta-m\left(\frac{p_{i_1}+\Delta}{p_{i_1}+m+\Delta}+\frac{p_{i_2}-\Delta}{p_{i_2}+m-\Delta}\right).
\end{aligned}$$

Now observe that

$$\begin{aligned}
&\frac{d}{dx}\left(\frac{p_{i_1}+x}{p_{i_1}+m+x}+\frac{p_{i_2}-x}{p_{i_2}+m-x}\right)\\
&\quad = \frac{d}{dx}\left(1-\frac{m}{p_{i_1}+m+x}+1-\frac{m}{p_{i_2}+m-x}\right)\\
&\quad = \frac{m}{(p_{i_1}+m+x)^2}-\frac{m}{(p_{i_2}+m-x)^2}\\
&\quad \leq 0
\end{aligned}$$

if $p_{i_1} \geq p_{i_2}$ and $0 \leq x < p_{i_2}$. As $m \geq 0$, we may conclude

$$p_{i_1}+p_{i_2}-m\left(\frac{p_{i_1}+\Delta}{p_{i_1}+m+\Delta}+\frac{p_{i_2}-\Delta}{p_{i_2}+m-\Delta}\right)$$

https://doi.org/10.1017/S0269964899133084 Published online by Cambridge University Press

is nondecreasing in Δ for $\Delta \in [0, p_{i_2})$. Thus,

$$(p_{i_1} + \Delta)^2 + (p_{i_1} + \Delta)^2 \left(\frac{p_{i_2} - \Delta}{1 - p_{i_2} + \Delta} \right) + (p_{i_2} - \Delta)^2 + (p_{i_2} - \Delta)^2 \left(\frac{p_{i_1} + \Delta}{1 - p_{i_1} - \Delta} \right) \geq p_{i_1}^2 + p_{i_1}^2 \frac{p_{i_2}}{1 - p_{i_2}} + p_{i_2}^2 + p_{i_2}^2 \frac{p_{i_1}}{1 - p_{i_1}}.$$

Putting all the terms together gives us the fact that $P(S(\infty) \leq 1)$ is Schur convex. ■

3. COUNTEREXAMPLE TO THE CONJECTURE

In this section, we will disprove the conjecture of Fill and Holst [2], hence show Theorem 2. Let $f(a,b,c,d,e) = P(S(\infty) \geq 4)$ if $n = 5$, $p_1 = a$, $p_2 = b$, $p_3 = c$, $p_4 = d$, and $p_5 = e$. Note that, if there are 5 objects, then $S(\infty)$ is at most 4. Thus, here $P(S(\infty) \geq 4) = P(S(\infty) = 4)$.

Note $f(0.36, 0.34, 0.20, 0.07, 0.03) \leq f(0.35, 0.35, 0.20, 0.07, 0.03)$ if the conjecture of Fill and Holst [2] holds. In Figure 1, Maple output shows this is not true. Via repeated procedure calls, the procedure five( ) goes through all 5! possible orderings and finds the probability that the ordering occurs in the stationary distribution multiplied by the probability that the last object in the ordering is picked. In the procedure one( ), the objects have probabilities (from top to bottom) b, c, d, e, and a. The Maple output also gives $(d^2/dx^2) f(0.35 + x, 0.35 - x, 0.20, 0.07, 0.03)$ at $x = 0$. The fact that this second derivative is positive at this point, combined with the fact that the first derivative is 0 at this point (by symmetry), gives a local minimum at $x = 0$ and not the local maximum required for Schur concavity.

Figure 2 gives a plot of $f(0.49 + x, 0.49 - x, 0.01, 0.005, 0.005)$. Note that $x = 0$ gives a local minimum here as well.

4. PROOF OF WHERE $P(S(\infty) \geq c)$ IS MAXIMIZED

In this section, we prove Theorem 3. Without loss of generality, we may suppose $c \in \{0, 1, \ldots, n\}$. Recall that the convention is that $S(\infty)$ is 0 if the top object is chosen. Thus, $S(\infty) \geq c$ precisely when an object not in the top c objects is picked.

Suppose each of the n objects is equally likely to be picked. Then the probability that the object is not in the top c objects is $(n - c)/n$ and

$$P(S(\infty) \geq c) = \sum_{i=1}^{n} \frac{1}{n} \frac{n - c}{n} = \frac{n - c}{n}.$$

Now, suppose the probability of picking each of the n objects is $p_1, p_2, \ldots, p_n$ (in order) with $p_1 \geq p_2 \geq \cdots \geq p_n > 0$ and $\sum_{i=1}^{n} p_i = 1$. Let e_i be the probability in the

https://doi.org/10.1017/S0269964899133084 Published online by Cambridge University Press

```
> ONE:=proc(a,b,c,d,e) a*b*c*d*e/((1-b)*(1-b-c)*(1-b-c-d)) end;

ONE := proc(a, b, c, d, e)
    a*b*c*d*e/((1 - b)*(1 - b - c)*(1 - b - c - d))
end

> TWO:=proc(a,b,c,d,e) ONE(a,b,c,d,e)+ONE(a,b,c,e,d) end;

TWO :=

    proc(a, b, c, d, e) ONE(a, b, c, d, e) + ONE(a, b, c, e, d) end

> THREE:=proc(a,b,c,d,e) TWO(a,b,c,d,e)+TWO(a,b,d,c,e)+TWO(a,b,e,c,d) end;

THREE := proc(a, b, c, d, e)
    TWO(a, b, c, d, e) + TWO(a, b, d, c, e) + TWO(a, b, e, c, d)
end

> FOUR:=proc(a,b,c,d,e) THREE(a,b,c,d,e)+THREE(a,c,b,d,e)+THREE(a,d,b,c,e)+THREE
(a,e,b,c,d) end;

FOUR := proc(a, b, c, d, e)
    THREE(a, b, c, d, e) + THREE(a, c, b, d, e)
     + THREE(a, d, b, c, e) + THREE(a, e, b, c, d)
end

> FIVE:=proc(a,b,c,d,e) FOUR(a,b,c,d,e)+FOUR(b,a,c,d,e)+FOUR(c,a,b,d,e)+FOUR(d,a
,b,c,e)+FOUR(e,a,b,c,d) end;

FIVE := proc(a, b, c, d, e)
    FOUR(a, b, c, d, e) + FOUR(b, a, c, d, e)
     + FOUR(c, a, b, d, e) + FOUR(d, a, b, c, e)
     + FOUR(e, a, b, c, d)
end

> FIVE(36/100,34/100,3/100,7/100,20/100);

                       30089275786008308O8589
                       -----------------------
                       515263795582977195O4000

> FIVE(35/100,35/100,3/100,7/100,20/100);

                          6272105125255319
                         ------------------
                         107407726684830000

> FIVE(36/100,34/100,3/100,7/100,20/100)-FIVE(35/100,35/100,3/100,7/100,20/100);

                      1167002824010990922 47
                     ---------------------------
                     2017257759707355718581600O0

> subs(x=0,diff(FIVE(35/100+x,35/100-x,3/100,7/100,20/100),x$2));

                       98833088184244738494O533
                      --------------------------
                      84932999840843355246600555

>
```

FIGURE 1. Maple output giving the counterexample.

https://doi.org/10.1017/S0269964899133084 Published online by Cambridge University Press

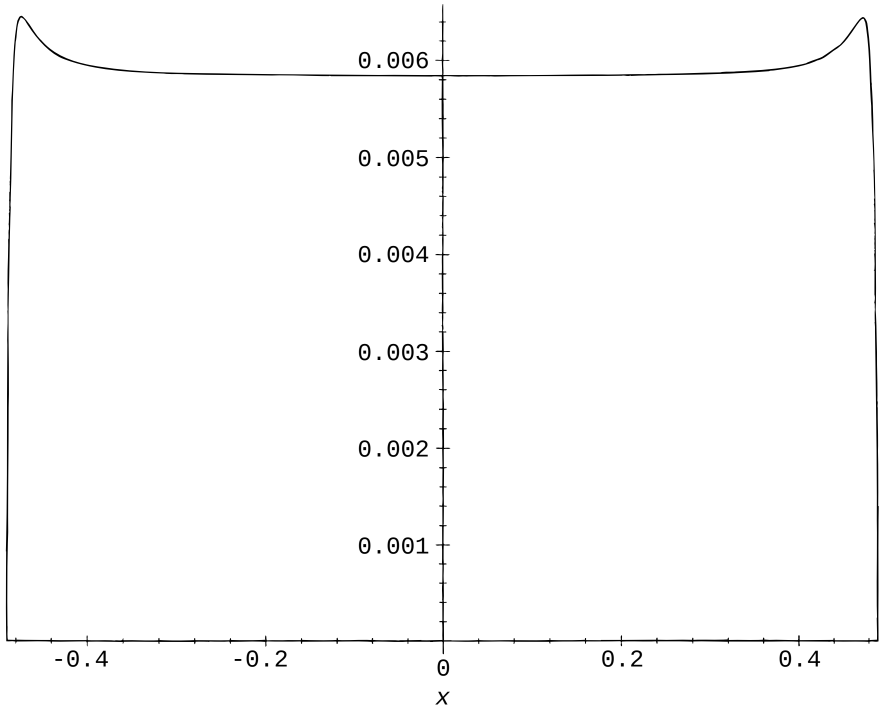

FIGURE 2. Maple plot of five$(0.49 + x, 0.49 - x, 0.01, 0.005, 0.005)$.

stationary distribution that object i is not in the top c objects. Note that e_i may depend on all of $p_1, p_2, \ldots, p_n$ as well as the choice of c.

PROPOSITION 4:

$$\sum_{i=1}^{n} e_i = n - c.$$

PROOF: This follows since the expected number of objects not in the top c is $n - c$. ∎

LEMMA 5: *If* $p_1 \geq p_2 \geq \cdots \geq p_n$, *then* $e_1 \leq e_2 \leq \cdots \leq e_n$.

PROOF: Suppose we perform i.i.d. trials with possible outcomes $1, \ldots, n$ such that outcome k has probability p_k. Suppose $i < j$; hence $p_i \geq p_j$. Use these trials to determine the orderings of two lists as follows. Suppose the lists are initially identical and suppose the initial distribution of a list is according to the stationary distribution of the move-to-front rule. Then for each trial use the outcome to reorder both lists as follows.

In list 1, if the outcome of the trial is k, then object k is selected and moved to the front.

https://doi.org/10.1017/S0269964899133084 Published online by Cambridge University Press

In list 2, if the outcome of the trial is k with $k \notin \{i, j\}$, then object k is selected. If the outcome is j, then object i is selected. If the outcome is i, then object j is selected with probability p_j/p_i and object i is selected otherwise. The selected object is moved to the top.

Note that, if object j is among the top c objects in list 1, then either object i is among the top c objects on list 2 or object j has not yet been selected in list 1. Since both lists move according to the move-to-front rule, after m steps we get that $(1 - e_j) \leq (1 - e_i) + P(S(j, m))$ where $S(j, m)$ is the event that object j has not been selected in list 1 in the first m steps. Since $\lim_{m\to\infty} P(S(j, m)) = 0$, we get $e_i \leq e_j$. ■

Lemma 5 holds for many reordering algorithms; one such algorithm is the move-ahead-one rule, where the selected object is moved ahead one in the list (except the object stays put if it is already on top).

To prove Theorem 3, define random variables I and J such that $P(I = i) = p_i$ and $P(J = i) = 1/n$ for $i = 1, \ldots, n$. Since p_i is nonincreasing, I is stochastically smaller than J. Thus,

$$\begin{aligned} P(S(\infty) \geq c) &= \sum_{i=1}^{n} p_i e_i \\ &= E[e_I] \\ &\leq E[e_J] \\ &= \sum_{i=1}^{n} \frac{e_i}{n} \\ &= \frac{n - c}{n} \end{aligned}$$

where $E[e_I] \leq E[e_J]$ by Lemma 5 and the stochastic ordering. Thus the theorem is proved.

5. QUESTIONS FOR FURTHER STUDY

It seems reasonable to believe that the conjecture of Fill and Holst [2] will also fail for larger values of c beyond those considered here, but a formal proof is not yet known.

A question worth exploring is the extent to which Schur concavity and Schur convexity fails. Figure 2 illustrates this failure, but the local maxima in the figure were only a few percent larger than the local minima at $x = 0$. In particular, consider $f(\vec{p}_1)/f(\vec{p}_2)$, where $f(\vec{p})$ is $P(S(\infty) \geq c)$ with probabilities determined by $\vec{p}$. What is the maximum of $f(\vec{p}_1)/f(\vec{p}_2)$ over all $\vec{p}_1$ and $\vec{p}_2$ such that any nonnegative Schur concave function g has $g(\vec{p}_1) \leq g(\vec{p}_2)$? Is there a bound on this maximum which works uniformly for all c? If so, what is the bound? If not, does each value of c have bound on this maximum?

Another question worth exploring is to see whether there are real-life examples where this lack of Schur concavity has a practical impact in slowing down a cache unexpectedly.

Acknowledgments

The author thanks Jim Fill for verifying a counterexample and for encouragement and suggestions. The author also thanks the reviewer for some suggestions.

References

1. Fill, J. (1998). Personal communication.
2. Fill, J. & Holst, L. (1996). On the distribution of search cost in the move-to-front rule. *Random Structures and Algorithms* 8: 179–186.
3. Hendricks, W.J. (1972). The stationary distribution of an interesting Markov chain. *Journal of Applied Probability* 9: 231–233.
4. Marshall, A. & Olkin, I. (1979). *Inequalities: Theory of majorization and its applications.* New York: Academic Press.