Properties of the miss ratio for a 2-level storage model with LRU or FIFO replacement strategy and Independent References

J. van den Berg [1] and **D. Towsley** [2]

Abstract

We consider the behavior of a 2-level storage system operating with the LRU or the FIFO replacement strategy where accesses to the main storage are described by the independent reference model (IRM). Let the size of main storage be m. We prove that the miss ratio (i.e. the steady-state probability that the item currently required is not in main storage) exhibits the following properties:

- The miss ratio is a convex function of m under LRU,
- The following function of the miss ratio, $(1 - \text{miss ratio})/m$, is non-increasing in m under FIFO,
- The miss ratio is a Schur-convex function of the reference probabilities under FIFO.

Last, we disprove some results claimed earlier in the literature.

[1] CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

[2] Dept. of Computer & Information Science, Univ. Massachusetts, Amherst, MA 01003,U.S.A. The work of this author was supported in part by the National Science Foundation under grants IRI-8908693 and DCR-8500332.

Key words. FIFO, independent reference model, LRU, memory management, miss ratio

1 Introduction

Suppose there are n items denoted by $1, 2, ..., n$ which are permanently located in auxiliary storage. In addition, $m \leq n$ items may also be in main storage. If an item is required, first the main storage is inspected. If it is not present there, it is copied from auxiliary storage into main storage. If the main storage is already full, an item is removed according to some replacement algorithm. This situation arises in many contexts, such as processor caches [15], paged virtual memory systems [7], database buffers [13], etc...

Many replacement algorithms have been studied in the past for such systems. Two of the most widely studied algorithms include the Least Recently Used (LRU) and First-In-First-Out (FIFO) algorithms. There is a large amount of literature concerned with analyzing the performance of these policies. The reader is referred to [1, chapters 4,5]. for a thorough treatment of this subject.

An important performance metric is the miss ratio, which is defined as the steady-state probability that the item currently required is not in main storage (in computer storage terms, this is the frequency of page faults). Much of the work in the area of 2-level storage systems is concerned with predicting this metric, designing practical algortihms for minimizing this metric, or choosing the size of the main storage so as to minimize the miss ratio subject to cost constraints. Unfortunately these tasks are made difficult due to the absence of simple to compute closed form expressions for this metric, even under the simplest assumptions. For example, the simplest model is one where each item has a fixed reference probability and where the references are mutually independent. In other words, the string of page references is modeled as a sequence of i.i.d. random variables. This is known as the independent reference model (IRM). In the literature explicit formulas for the miss ratio in terms of the reference probabilities and the size of main storage are given (see [12, 8] for examples). However, they involve a summation in which the number of terms grows exponentially as a function of m, and are therefore not very tractable. A number of papers have focussed on developing bounds (see [10, 2] for examples) or developing approximate models, [6], for predicting the miss ratio.

In this paper we derive several properties for the miss ratio under the LRU and FIFO policies for the IRM. These include,

- The miss ratio is a convex function of m under LRU,
- The following function of the miss ratio, $(1 - \text{miss ratio})/m$, is non-increasing in m under FIFO,

- The miss ratio is a Schur-convex function of the reference probabilities under FIFO.

In addition, we construct a counterexample to some claims made regarding the Schur-convexity of the miss ratio under LRU in [1].

The above properties are of interest for several reasons. First, they can be used by analysts to develop accurate approximate models for predicting the miss ratio of LRU and FIFO, i.e., these approximate models should satisfy the above properties. Second, a number of studies, [16, 6], have studied the problem of partitioning a main storage operating under LRU among several classes of items in order to minimize the miss ratio. In all of these studies, the approaches require that the miss ratio be a convex function of m in order to guarantee that the optimal solution be found. Our result provides evidence that this is so.

Last, although our results are for the IRM, they are of practical interest because there is evidence that the IRM is appropriate for database systems, [11], and this has formed the basis of a number of analytical models of such systems operating under LRU, [4, 5].

2 Formal statement of results

Consider the IRM model described in section 1. Let n denote the number of items, $\boldsymbol{p} = (p_1, p_2, ..., p_n)$ their reference probabilities, and m the size of main storage. The miss ratio, introduced before, depends on $\boldsymbol{p}$, m and the replacement algorithm A, and is denoted by $F_m(A, \boldsymbol{p})$. Often, when the context is clear, we will omit one or more of the parameters.

Let $\{x_k\}_{k=0}^{\infty}$ be a sequence of independent and identically distributed random variables with

$$p_j = P(x_0 = j), \quad j = 1, 2, \ldots, n, \tag{1}$$

where x_0 can be regarded as the item currently being referenced and x_k the k-th item previously referenced, $k = 1, 2, \ldots$. It is not difficult to see (see e.g. Aven, Coffman and Kogan [1, chapter 4]) that for LRU the miss ratio, defined in the introduction, is equal to:

$$\begin{aligned} F_m(\text{LRU}) \quad = \quad & P(x_0 \text{ is not equal to one of the first } m \text{ distinct values} \\ & \text{in the sequence } x_1, x_2, \ldots). \end{aligned} \tag{2}$$

Using this equality, the miss ratio can also be expressed more explicitly in terms of the p_i's and m (see [12, 8]). We will find (2) suitable for our purpose.

As to FIFO, it has been shown by King [12] and Aven and Sokolov [3] that, for $m \leq n$,

$$F_m(\text{FIFO}) = \frac{\sum_{(i_1,\ldots,i_m)\in\Lambda_{m,n}} p_{i_1}p_{i_2}\cdots p_{i_m}(1 - p_{i_1} - p_{i_2} \cdots - p_{i_m})}{\sum_{(i_1,\ldots,i_m)\in\Lambda_{m,n}} p_{i_1}p_{i_2}\cdots p_{i_m}}, \tag{3}$$

where the sum is over the set $\Lambda_{m,n}$ of all sequences of m distinct elements taken from the set of integers $\{1, 2, \ldots, n\}$ (see also [1, section 4.6]). If $m > n$ then, of course, F_m is 0. It is easily seen that (3) can also be written as follows:

$$F_m(\text{FIFO}) = \frac{P(x_0, x_1, \ldots, x_m \text{ are distinct})}{P(x_1, \ldots, x_m \text{ are distinct})}, \tag{4}$$

and this is the expression used in the remainder of this paper.

We will prove the following two results:

Theorem 1 *The miss ratio for LRU is a convex function of the size of main storage, i.e. for fixed* $\boldsymbol{p}$, $F_m(LRU) - F_{m+1}(LRU)$ *is non-increasing in* m.

Theorem 2 *Under FIFO, the miss ratio exhibits the following properties:*

a) For a fixed $\boldsymbol{p}$, $(1 - F_m(FIFO))/m$ *is non-increasing in* m.

b) The miss ratio for FIFO is a Schur-convex function of the reference probability vector $\boldsymbol{p}$.

As a consequence of theorem 2 b), we have the following corollary.

Corollary 1 *Let* $V \subset \{1, \ldots, n\}$ *and* $0 \leq \alpha \leq 1$. *Then, subject to the constraint* $\sum_{i\in V} p_i = \alpha$, $\max_{\boldsymbol{p}} F_m(FIFO, \boldsymbol{p})$ *is achieved for* $p_i = \frac{\alpha}{|V|}$, $i \in V$ *and* $p_i = \frac{1-\alpha}{n-|V|}$, $i \notin V$.

Remark. The following special case of theorem 2 b) has been stated in [1, sec. 4.7]. If $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{p}' = (p'_1, \ldots, p'_n)$ are two memory reference probability distributions such that $p_1 \geq p_2 \geq \cdots \geq p_n$, $p'_1 \geq p'_2 \geq \cdots \geq p'_n$, $p_i = p'_i$, $i = 3, \ldots, n$, $p_1 + p_2 = p'_1 + p'_2$, and $p'_1 \geq p_1$, then $F_m(\text{FIFO}, \boldsymbol{p}) \geq F_m(\text{FIFO}, \boldsymbol{p}')$. As a consequence they have obtained the above corollary under the additional restriction that for any $i \in V$ and $j \notin V$ $p_i \geq p_j$. They claim that their results hold for a large class of algorithms, including LRU and then develop upper bounds on the miss ratio of LRU based on this claim. However, we have constructed a counter-example for LRU (see section 5).

In section 3 we prove theorem 1. In section 4 we prove theorem 2. The corollary above follows straightforwardly from the second property of theorem 2 and we omit its proof. In section 5 we construct a counter-example to some results claimed in [1].

3 Convexity of LRU

Since this section deals with LRU only, we omit this argument in the mathematical expressions. Let $g(\boldsymbol{p}, m) = F_{m-1}(\boldsymbol{p}) - F_m(\boldsymbol{p})$. From (2) it follows:

$$\begin{aligned} g(\boldsymbol{p}, m) &= P(x_0 \text{ is equal to the } m\text{-th different value} \\ &\qquad \text{in the sequence } x_1, x_2, \ldots). \end{aligned} \tag{5}$$

It suffices to prove that $g(\boldsymbol{p}, m)$ is a non-increasing function of m, $m = 1, 2, \ldots$. We prove this using induction.

Consider $m = 1$ as the basis step. Note that

$$g(\boldsymbol{p}, 1) = P(x_0 = x_1) = \sum_{i=1}^{n} {p_i}^2 \tag{6}$$

and

$$\begin{aligned} g(\boldsymbol{p}, 2) &= P(x_0 \text{ is equal to the second different value in the sequence} \\ &\qquad x_1, x_2, \ldots) \\ &= \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} P(x_0 = i, x_1 = j, \text{ the first value in the sequence} \\ &\qquad x_2, x_3, \ldots \text{ different from } j \text{ equals } i) \\ &= \sum_{i=1}^{n} \left(\sum_{j=1, j\neq i}^{n} p_i p_j \frac{p_i}{1 - p_j} \right). \end{aligned} \tag{7}$$

Now define, for $1 \le i \le n$, $f_i = p_i$ and $h_i = \sum_{j=1, j\neq i}^{n} \frac{p_j}{1-p_j}$.

Equation (7) can then be written as

$$g(\boldsymbol{p}, 2) = \sum_{i=1}^{n} p_i f_i h_i. \tag{8}$$

It is easy to check that for any i and j, $1 \le i, j \le n$,

$$f_i \ge f_j \iff h_i \le h_j. \tag{9}$$

A degenerate case of the FKG inequality (Fortuin, Kasteleyn and Ginibre [9]) states that (9) implies

$$\sum_{i=1}^{n} p_i f_i h_i \leq (\sum_{i=1}^{n} p_i f_i)(\sum_{i=1}^{n} p_i h_i). \tag{10}$$

Remark: We speak of a *degenerate* case of the FKG inequality because (9) involves a total order, while the *general* FKG inequality holds, under certain conditions, for partial orders as well. The proof of the special case above is easy (see the introduction of [9]).

From (8), (10) and the definition of f_i and g_i we get

$$\begin{aligned} g(\boldsymbol{p}, 2) &\leq (\sum_{i=1}^{n} p_i f_i)(\sum_{i=1}^{n} p_i h_i) \\ &= (\sum_{i=1}^{n} {p_i}^2)(\sum_{i=1}^{n} p_i \sum_{j=1, j\neq i}^{n} \frac{p_j}{1-p_j}) \\ &= g(\boldsymbol{p}, 1) \sum_{j=1}^{n} \frac{p_j}{1-p_j} \sum_{i=1, i\neq j}^{n} p_i \\ &= g(\boldsymbol{p}, 1) \sum_{j=1}^{n} \frac{p_j}{1-p_j}(1-p_j) \\ &= g(\boldsymbol{p}, 1). \end{aligned} \tag{11}$$

This completes the proof for $m = 1$.

Now let $m \geq 2$ and assume the result holds for $m-1$. We have

$$\begin{aligned} g(\boldsymbol{p}, m) &= P(x_0 \text{ is the } m\text{-th different value in } x_1, x_2, \ldots), \\ &= \sum_{j=1}^{n} P(x_1 = j, x_0 \neq j, x_0 \text{ is the } (m-1)\text{-th different value in} \\ &\qquad {x_1}^{(j)}, {x_2}^{(j)}, \ldots) \end{aligned} \tag{12}$$

where ${x_1}^{(j)}, {x_2}^{(j)}, \ldots$ is the sequence obtained from $x_1, x_2, \ldots$ by removing all elements with value j. It follows from standard arguments that, given $x_1 = j$ and $x_0 \neq j$, the sequence $x_0, {x_1}^{(j)}, {x_2}^{(j)}, \ldots$ is a sequence of i.i.d. random variables with probability distribution given by the vector

$$\boldsymbol{p}^{(j)} \equiv (\frac{p_1}{1-p_j}, \frac{p_2}{1-p_j}, \ldots, \frac{p_{j-1}}{1-p_j}, \frac{p_{j+1}}{1-p_j}, \ldots, \frac{p_n}{1-p_j}). \tag{13}$$

Hence, (12) can be rewritten as follows:

$$g(\boldsymbol{p}, m) = \sum_{j=1}^{n} p_j(1-p_j) g(\boldsymbol{p}^{(j)}, m-1). \tag{14}$$

Of course $g(\boldsymbol{p}, m+1)$ can be written in a similar way, the only difference being that the $m-1$ in the r.h.s. of (14) is replaced by m. Now apply the induction hypothesis. ∎

4 Properties of FIFO

4.1 Notation and a preliminary result

Since this section deals with FIFO only, we omit that argument in the mathematical expressions. First we need some additional notation. Let the sequence $x_0, x_1, \ldots$ be as in section 2. Define, for $k = 1, 2, \ldots$,

$$G_k(\boldsymbol{p}) = P(x_1, \ldots, x_k \text{ are distinct}). \tag{15}$$

Moreover, define, for $V \subset \{1, \ldots, n\}$,

$$G_k^V(\boldsymbol{p}) = P(x_1, \ldots, x_k \text{ are distinct and not in } V). \tag{16}$$

We will often omit the parameter $\boldsymbol{p}$ when it does not introduce any confusion. If $k \leq 0$, we define $G_k^V = 1$. Clearly, if $V = \emptyset$, then $G_k^V = G_k$. Also note that, if $i \notin V$ and $k \geq 0$, then

$$\begin{aligned} G_k^V &= P(x_1, \ldots, x_k \text{ are distinct and not in } V) \\ &= P(x_1, \ldots, x_k \text{ are distinct and not in } V \cup \{i\}) \\ &\quad + P(x_1, \ldots, x_k \text{ are distinct and not in } V, \text{ and} \\ &\qquad \text{exactly one of them equals } i) \\ &= G_k^{V \cup \{i\}} + k p_i G_{k-1}^{V \cup \{i\}}. \end{aligned} \tag{17}$$

We require the following lemma.

Lemma 1 *Let* $1 \leq k \leq l$ *and* $V \subset \{1, \ldots, n\}$. *Then*

$$G_k^V G_l^V \geq G_{k-1}^V G_{l+1}^V. \tag{18}$$

Proof. If $G^V_{l+1} = 0$, (18) is clearly true. If $G^V_{l+1} \neq 0$ then $G^V_i \neq 0$ for all $i \leq l+1$ and (18) can be written as $G^V_k / G^V_{k-1} \geq G^V_{l+1}/G^V_l$ so that it is sufficient to handle the case $k = l$. So we are required to prove

$$(G^V_k)^2 - G^V_{k-1}G^V_{k+1} \geq 0. \tag{19}$$

This can be done by induction as follows. The case $k \leq 1$ is trivial. Now let $k > 1$ and assume that the inequality holds for $k-1$. We have

$$\begin{aligned}(G^V_k)^2 - G^V_{k-1}G^V_{k+1} \quad = \quad & P(x_1, \ldots, x_k \text{ are distinct and not in V,} \\ & \qquad x_{k+1}, \ldots x_{2k} \text{ are distinct and not in V)} \\ & - P(x_1, \ldots, x_{k-1} \text{ are distinct and not in V,} \\ & \qquad x_k, \ldots, x_{2k} \text{ are distinct and not in } V).\end{aligned}$$

Since for any two events B and C we have $P(B) - P(C) = P(B \cap C^c) - P(C \cap B^c)$, the above can be rewritten as

$$\begin{aligned}&(G^V_k)^2 - G^V_{k-1}G^V_{k+1} = \\ &\quad P(x_1, \ldots, x_k \text{ are distinct and not in } V, x_{k+1}, \ldots, x_{2k} \text{ are distinct} \\ &\qquad\qquad \text{and not in } V, x_k, \ldots, x_{2k} \text{ are not distinct)} \\ &\quad - P(x_1, \ldots, x_{k-1} \text{ are distinct and not in } V, x_k, \ldots, x_{2k} \text{ are distinct} \\ &\qquad\qquad \text{and not in } V, x_1, \ldots, x_k \text{ are not distinct).}\end{aligned} \tag{20}$$

The first event occurs if and only if, for some $i \in \{1, \ldots, n\} - V$, and $l \in \{k+1, \ldots, 2k\}$, $x_1, \ldots x_{k-1}$ are not in $V \cup \{i\}$, $x_k = x_l = i$ and $x_{k+1}, \ldots, x_{l-1}, x_{l+1}, \ldots, x_{2k}$ are distinct and not in $V \cup \{i\}$. This occurs with probability

$$\sum_{i \in \{1,\ldots,n\} \cap V^c} k{p_i}^2 (G^{V \cup \{i\}}_{k-1})^2.$$

Similarly, the second event has probability

$$\sum_{i \in \{1,\ldots,n\} \cap V^c} (k-1){p_i}^2 G^{V \cup \{i\}}_{k-2} G^{V \cup \{i\}}_k.$$

We have

$$(G^V_k)^2 - G^V_{k-1}G^V_{k+1} \quad = \quad \sum_{i \in \{1,\ldots,n\} \cap V^c} \left[k{p_i}^2 (G^{V \cup \{i\}}_{k-1})^2 - (k-1){p_i}^2 G^{V \cup \{i\}}_{k-2} G^{V \cup \{i\}}_k\right]$$

$$\geq \sum_{i\in\{1,\ldots,n\}\cap V^c} (k-1)\left[{p_i}^2(G_{k-1}^{V\cup\{i\}})^2 - {p_i}^2 G_{k-2}^{V\cup\{i\}} G_k^{V\cup\{i\}}\right]$$

$$\geq 0.$$

This last inequality is a consequence of the induction hypothesis. ■

4.2 Proof of theorem 2a

We have to prove $(1-F_m)/m \geq (1-F_{m+1})/(m+1)$. If $m+1>n$, then F_m and F_{m+1} are 0 and the inequality is obviously true. So we may assume $m+1\leq n$. From (4) it follows that

$$\begin{aligned}\frac{1-F_m}{m} &= \frac{1}{m}\times\frac{P(x_1,\ldots x_m \text{ are distinct and one of them equals } x_0)}{P(x_1,\ldots,x_m \text{ are distinct})},\\ &= \frac{P(x_0=x_1 \text{ and } x_1,\ldots x_m \text{ are distinct})}{P(x_1,\ldots,x_m \text{ are distinct})}.\end{aligned}$$

and, similarly,

$$\frac{1-F_{m+1}}{m+1} = \frac{P(x_0=x_1 \text{ and } x_1,\ldots x_{m+1} \text{ are distinct})}{P(x_1,\ldots,x_{m+1} \text{ are distinct})}.$$

Hence, we have to prove

$$\begin{aligned}&P(x_0=x_1 \text{ and } x_1,\ldots,x_m \text{ are distinct})\\ &\times P(x_1,\ldots,x_{m+1} \text{ are distinct})\\ &\geq P(x_0=x_1 \text{ and } x_1,\ldots,x_{m+1} \text{are distinct})\\ &\quad\times P(x_1,\ldots,x_m \text{ are distinct}).\end{aligned}\tag{21}$$

By summing over all possible values of x_1, the first factor in the l.h.s. of (21) equals

$$\sum_{i=1}^{n} P(x_0=x_1=i)P(x_2,\ldots,x_m \text{ are distinct and different from } i) =$$

$$\sum_{i=1}^{n} p_i^2 G_{m-1}^{\{i\}},$$

while the second factor equals

$$\sum_{j=1}^{n} P(x_1 = j)P(x_2, \ldots, x_{m+1} \text{ are distinct and different from } j) =$$

$$\sum_{j=1}^{n} p_j G_m^{\{j\}}.$$

The two factors in the r.h.s. can be rewritten similarly and we conclude that (21) is equivalent to

$$\sum_{1 \le i,j \le n} p_i^2 p_j G_{m-1}^{\{i\}} G_m^{\{j\}} \ge \sum_{1 \le i,j \le n} p_i^2 p_j G_m^{\{i\}} G_{m-1}^{\{j\}}$$

or

$$\sum_{1 \le i < j \le n} (p_i - p_j) p_i p_j \left[G_{m-1}^{\{i\}} G_m^{\{j\}} - G_{m-1}^{\{j\}} G_m^{\{i\}} \right] \ge 0.$$

This last inequality is certainly true if, for any pair $i, j \in \{1, \ldots, n\}$,

$$p_i \ge p_j \iff G_{m-1}^{\{i\}} G_m^{\{j\}} - G_{m-1}^{\{j\}} G_m^{\{i\}} \ge 0. \tag{22}$$

In order to see that (22) is true, note that, if $i \neq j$, by (17),

$$G_{m-1}^{\{i\}} G_m^{\{j\}} = (G_{m-1}^{\{i,j\}} + (m-1) p_j G_{m-2}^{\{i,j\}})(G_m^{\{i,j\}} + m p_i G_{m-1}^{\{i,j\}}),$$

and the expression for $G_{m-1}^{\{j\}} G_m^{\{i\}}$ is obtained by exchanging i and j. Hence,

$$G_{m-1}^{\{i\}} G_m^{\{j\}} - G_{m-1}^{\{j\}} G_m^{\{i\}} =$$

$$m(p_i - p_j) \left(G_{m-1}^{\{i,j\}} \right)^2 - (m-1)(p_i - p_j) G_{m-2}^{\{i,j\}} G_m^{\{i,j\}}.$$

According to lemma 1 this is non-negative iff $p_i \ge p_j$, so that (22) is indeed true. ∎

4.3 Proof of theorem 2b

We begin this section with the definition of a Schur-convex function.

Definition 1 *Vector* $\boldsymbol{X} = (X_1, \cdots, X_k)$ *is said to majorize vector* $\boldsymbol{Y} = (Y_1, \cdots, Y_k)$ *(written* $\boldsymbol{Y} \prec \boldsymbol{X}$*) iff*

$$\sum_{i=1}^{j} \hat{Y}_i \leq \sum_{i=1}^{j} \hat{X}_i, \quad j = 1, \cdots, k-1$$

$$\sum_{i=1}^{k} \hat{Y}_i = \sum_{i=1}^{k} \hat{X}_i$$

where the notation $\hat{X}_i$ *(resp.* $\hat{Y}_i$*) is taken to be the i-th largest element of* $\boldsymbol{X}$ *(resp.* $\boldsymbol{X}$*).*

Definition 2 *A real valued function* f *defined on* $I\!R^k$ *is said to be Schur-convex if*

$$\boldsymbol{x} \prec \boldsymbol{y} \quad \Rightarrow \quad f(x) \leq f(y), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in I\!R^k$$

We find the following result useful in our proof.

Lemma 2 (Marshall and Olkin [14, section 3.A.2.b]) *A function* f *is Schur-convex iff* f *is symmetric and* $f(\lambda q, (1-\lambda)q, y_3, y_4, \ldots, y_k)$ *is a non-decreasing function of* λ *for* $\lambda \in (0, 1/2]$.

In our problem, F_m is clearly a symmetric function of $\boldsymbol{p}$. Hence it suffices to fix $p_3, \ldots, p_n$, define $p_1 = \lambda q$ and $p_2 = (1-\lambda)q$ for some $\lambda \in (0,1)$ where $q = 1 - (p_3 + \ldots + p_n)$ and show that F_m is non-decreasing in λ for $\lambda \in (0, 1/2]$. To avoid trivialities we assume that $1 \leq m < n$.

Note that, by (4) and (15),

$$F_m = G_{m+1}/G_m. \tag{23}$$

Further, by the same arguments used in the derivation of (17),

$$G_m = G_m^{\{1,2\}} + mqG_{m-1}^{\{1,2\}} + m(m-1)q^2\lambda(1-\lambda)G_{m-2}^{\{1,2\}}. \tag{24}$$

Hence,

$$\frac{d}{d\lambda}G_m = (1-2\lambda)q^2 m(m-1)G_{m-2}^{\{1,2\}}. \tag{25}$$

Of course, G_{m+1} and $\frac{d}{d\lambda}G_{m+1}$ are obtained from the above expressions by replacing m by $m+1$.

We have to prove that

$$\frac{d}{d\lambda}\frac{G_{m+1}}{G_m} \geq 0,$$

which is equivalent to

$$G_m \frac{d}{d\lambda} G_{m+1} \geq G_{m+1} \frac{d}{d\lambda} G_m,$$

which, by (24) and (25) and the fact that $\lambda \in (0, 1/2]$, is equivalent to

$$(G_m^{\{1,2\}} + mqG_{m-1}^{\{1,2\}} + m(m-1)q^2\lambda(1-\lambda)G_{m-2}^{\{1,2\}}) \times (m+1)G_{m-1}^{\{1,2\}}$$
$$\geq$$
$$(G_{m+1}^{\{1,2\}} + (m+1)qG_m^{\{1,2\}} + m(m+1)q^2\lambda(1-\lambda)G_{m-1}^{\{1,2\}}) \times (m-1)G_{m-2}^{\{1,2\}}.$$

After removing common terms from both sides, the above inequality reduces to

$$\begin{array}{rcl} (m+1)G_m^{\{1,2\}}G_{m-1}^{\{1,2\}} & + & m(m+1)q(G_{m-1}^{\{1,2\}})^2 \\ & \geq & \\ (m-1)G_{m+1}^{\{1,2\}}G_{m-2}^{\{1,2\}} & + & (m-1)(m+1)qG_{m-2}^{\{1,2\}}G_m^{\{1,2\}}, \end{array} \tag{26}$$

which follows from lemma 1. ∎

5 Counterexample

In [1], section 4.7, two properties, denoted by P1 and P2 are stated and used in the proof of several theorems. The claim is made that P1 and P2 "are satisfied by any reasonable algorithm" including LRU [3]. However, we present a counter-example to P2 for LRU. Before we state property P2 we introduce some additional notation.

If A is a replacement algorithm, then $R(1)A$ is defined to be the replacement algorithm corresponding to A, where item 1 resides permanently in main storage and algorithm A is applied with respect to the $m-1$ remaining positions in main storage and the $n-1$ remaining items.

It is not difficult to see that (see [1, sections 4.6 and 4.7]) the following relation is true for many algorithms, including LRU

$$F_m(R(1)A, \boldsymbol{p}) = (1-p_1)F_{m-1}(A, \boldsymbol{p}^{(1)}), \tag{27}$$

where $\boldsymbol{p}^{(1)} = (p_2/(1-p_1), \ldots, p_n/(1-p_1))$. Property P2 is stated below.

[3] This material originally came from [2] where no such claim was made.

Property P2. *Let* $\boldsymbol{p}$ *and* $\boldsymbol{p}'$ *be access distributions such that* $p_1 \geq p_2 \geq \ldots \geq p_n$, $p_1' \geq p_2' \geq p_3$, $p_1' + p_2' = p_1 + p_2$, $p_1' > p_1$, *and* $p_i' = p_i$, $i = 3, \ldots, n$. *Then*

$$F_m(R(1)A, \boldsymbol{p}') < F_m(R(1)A, \boldsymbol{p}) \tag{28}$$

whenever $m < n$.

We proceed with our counter-example. Let $\boldsymbol{p}$ and $\boldsymbol{p}'$ satisfy the condition in P2. Let $q = p_1 + p_2 = p_1' + p_2'$, and $p_3 = p_4 = \ldots = p_n = p_3' = p_4' = \ldots = p_n' = (1-q)/n$. Let $\{y_k\}_{k=0}^{\infty}$ be a sequence of i.i.d. random variables with distribution given by $\boldsymbol{p}^{(1)}$ (see (27)).

According to (2) we have, for $A = \text{LRU}$,

$$\begin{aligned} F_{m-1}(A, \boldsymbol{p}^{(1)}) &= P(y_0 = 2 \text{ and the first } m-1 \text{ distinct values in} \\ &\qquad y_1, y_2, \ldots \text{ are different from } 2) \\ &\quad + P(y_0 \neq 2 \text{ and } y_0 \text{ does not belong to the first } m-1 \\ &\qquad\quad \text{distinct values in } y_1, y_2, \ldots). \end{aligned} \tag{29}$$

Now fix all parameters, except n, which we allow to go to infinity (note that $\boldsymbol{p}$ and $\boldsymbol{p}'$ depend on n). As to the first term in (29), the event (the first $m-1$ distinct values in $y_1, y_2, \ldots$ are different from 2) is contained in the event ($y_1, y_2, \ldots, y_{m-1}$ are different from 2). However, the difference between the two events is contained in the event ($y_1, y_2, \ldots, y_{m-1}$ are different from 2, but they are not all distinct), whose probability goes to 0 as n goes to infinity.

As to the second term in (29), the difference between the event ($y_0 \neq 2$ and y_0 does not belong to the first $m-1$ distinct values in $y_1, y_2, \ldots$) and the event ($y_0 \neq 2$) is the event ($y_0 \neq 2$ and y_0 belongs to the first $m-1$ distinct values in $y_1, y_2, \ldots$), whose probability also goes to 0 as n goes to infinity.

Therefore, we have

$$\begin{aligned} \lim_{n\to\infty} F_{m-1}(A, \boldsymbol{p}^{(1)}) &= P(y_0 = 2 \text{ and } y_1, \ldots, y_{m-1} \neq 2) + P(y_0 \neq 2), \\ &= \frac{p_2}{1-p_1}\left(\frac{1-q}{1-p_1}\right)^{m-1} + \frac{1-q}{1-p_1}. \end{aligned}$$

Hence, by (27),

$$\lim_{n\to\infty} F_m(R(1)A, \boldsymbol{p}) = p_2\left(\frac{1-q}{1-p_1}\right)^{m-1} + 1 - q \tag{30}$$

and analogously,

$$\lim_{n\to\infty} F_m(R(1)A, \boldsymbol{p}') = p_2'(\frac{1-q}{1-p_1{}'})^{m-1} + 1 - q. \tag{31}$$

However, by the assumptions, $p_1{}' > p_1$, it follows that $\frac{1-q}{1-p_1} < \frac{1-q}{1-p_1{}'}$, so that, for m sufficiently large,

$$p_2(\frac{1-q}{1-p_1})^{m-1} < p_2{}'(\frac{1-q}{1-p_1{}'})^{m-1}.$$

Fix such an m. According to (30) and (31) there exists an n such that

$$F_m(R(1)A, \boldsymbol{p}) < F_m(R(1)A, \boldsymbol{p}'),$$

which contradicts P2.

Property P2 plays an important role in the derivation of other results in section 4.7 of [1]. For instance, in Theorem 14 it is stated that, for a large class of algorithms, including LRU, the following holds:

If $p_1 \geq p_2 \geq \cdots \geq p_n$ and $\sum_{i=s+1}^{n} p_i = \alpha$, then

$$F_m(A, \boldsymbol{p}) \leq F_m(A, \boldsymbol{p}')$$

where $p_1{}' = \ldots = p_s{}' = (1-\alpha)/s$ and $p_{s+1}{}' = \ldots = p_n{}' = \alpha/(n-s)$. Here A is taken to be one of these algorithms. (For $A =$ FIFO, this is a somewhat weaker version of corollary 1 in section 2, see the remark after our corollary). A small modification of the construction of our counterexample to P2 leads quite easily to a counterexample (for LRU) to the above stated theorem 14 in [1]. Theorem 14 in turn is used to develop several distribution free upper bounds on the miss ratio of LRU.

6 Summary

In this paper we have derived several interesting properties of the miss ratio for LRU and FIFO in a 2-level storage model for the IRM. Using similar arguments, it is possible to derive the convexity result for LRU to the transient case, i.e., the miss ratio for the i-th reference is a convex function of the main storage size m under the assumption that the main storage is empty prior to the first reference. Extending the properties for the miss ratio under FIFO to the transient case appears to be more difficult.

Another direction worth pursuing is that of obtaining similar results for more interesting reference models where the i-th reference may depend in some way on the previous $j > 0$ references (the IRM corresponds to this model with $j = 0$).

Acknowledgments

We thank Philippe Robert for useful discussions.

References

[1] Aven, O.I., Coffman, E.G., Jr., and Kogan, Y.A. *Stochastic Analysis of Computer Storage*, Reidel Publishing Company, 1987.

[2] Aven, O.I., Boguslavsky, L.B., Kogan, Y.A., "Some Results on Distribution-Free Analysis of Paging Algorithms," *IEEE Trans. on Computers*, **C-25**, 7, (1976), 737-745.

[3] Aven, O.I. and Sokolov, V.B., "On Some Methods of Memory Management", *Technical Cybernetics*, **2** (1971), 89-94 (In Russian).

[4] Dan, A., Dias, D.M., Yu, P.S., "The Effect of Skewed Data Access on Buffer Hits and Data Contention in a Data Sharing Environment," *Proc. 16-th International Conf. on Very Large Databases*, Brisbane, Australia, Aug. 1990, 419-431.

[5] Dan, A., Dias, D.M., Yu, P.S., "Analytical Modelling of a Hierarchical Buffer for a Data Sharing Environment," *Proc. 1991 ACM SIGMETRICS Conf.on Measurement and Modeling of Computer Systems*, San Diego, U.S.A., May 1991, 156-167.

[6] Dan, A. and Towsley, D., "An Approximate Analysis of the LRU and FIFO Buffer Replacement Schemes," *Proc. 1990 ACM SIGMETRICS Conf.on Measurement and Modeling of Computer Systems*, Boulder, Colorado, USA, May 1990, 143-152.

[7] Denning, P.J., "Virtual Memory," *Comput. Surv.*, **2** (1970), 153-189.

[8] Flajolet, P., Gardy, D., and Thimonier, L., "Birthday Paradox, Coupon Collectors, Caching Algorithms and Self Organizing Search", Tech. Rep. 720 INRIA, Rocquencourt, France, Aug. 1987.

[9] Fortuin, C.M., Kasteleyn, P.W. and Ginibre, J., "Correlation Inequalities On Some Partially Ordered Sets", *Communications in Mathematical Physics*, **22** (1971), 89-103.

[10] Franaszek, P.A. and Wagner, T.J., "Some Distribution-free Aspects of Paging Algorithm Performance," *J. Ass. Comput. Mach.*, **21**, (1974), 31-39.

[11] Kearns, J.P. and DeFazio, S., "Diversity in Database Reference Behavior," *Proc. 1989 ACM SIGMETRICS and PERFORMANCE'89 International Conf.on Measurement and Modeling of Computer Systems*, Berkeley, California, USA, May 1989, 11-19.

[12] King, W.F., "Analysis of Paging Algorithms", *Proc. IFIP Cong.*, Lyublyana, Yugoslavia, Aug. 1971, 485-490.

[13] Lang,T. and Fernandez, I.B., "Database Buffer Paging in Virtual Storage Systems," *ACM Trans. on Database Systems*, **2**, (1977), 339-351.

[14] Marshall, A.W., Olkin, I. *Inequalities: Theory of Majorization and its Applications*, Academic Press, 1979.

[15] Smith, A.J., "Cache Memories," *Comput. Surv.,* **14** (1982), 473-530.

[16] Stone, H.S., Wolf, J.L., and Turek, J., "Optimal Partitioning of Cache Memory," IBM Research Report RC14444 (64697), 1989. To appear in *ACM Trans. on Computer Systems.*