The Independence of Miss Ratio on Page Size

RONALD FAGIN AND MALCOLM C. EASTON

IBM Thomas J. Watson Research Center, Yorktown Heights, New York

ABSTRACT A theoretical justification is given to the empirical observation that in some computing systems with a paged, 2-level storage hierarchy, long-term miss ratio is roughly independent of page size Let *MISS* be the expected working-set miss ratio in the independent reference model, with expected working set size *CAP* pages Now form blocks, by combining the *B* pages with the highest probabilities of reference into one block, the *B* pages with the next-highest probabilities of reference into a second block, and so on Let *MISS** be the expected working-set miss ratio when all data are moved in blocks and when the expected working set size is again *CAP* pages, that is, $CAP/B = C$ blocks. It is proved that $|MISS - MISS^*| < (2/C) + (33/C^2)$. Thus, if the expected working-set size (in blocks) is sufficiently large, then the miss ratios in the blocked and unblocked cases are approximately equal This result is used to argue the approximate independence of miss ratio on page size in more realistic models of page references.

KEY WORDS AND PHRASES. page size, working set, LRU, miss ratio, storage hierarchy, independent reference model

CR CATEGORIES: 4 0, 4.3, 6.1, 6.20

1. *Introduction*

An important parameter in the design of a paged computing system is the page size, that is, the number of bytes of information transferred from one level of a storage hierarchy to another in case of a page fault. Among the factors which influence the choice of page size are the page fault rate (or "miss ratio"), the fragmentation of memory, and the access and transfer times of secondary memory devices (see Gelenbe et al. [12] for a more detailed discussion).

The research for this paper was sparked by an empirical observation of Bennett [3], who examined a page reference trace[1] from the IBM Advanced Administrative System (A.A.S.) [22], a large internal IBM data management system. Bennett found no consistent relationship between miss ratio and page size—for some main (first-level) memory sizes, the miss ratio was slightly larger for the larger page size, and for other main memory sizes, slightly smaller. In all cases the size of main memory had a vastly greater effect on miss ratio than did page size, if the page size was sufficiently large (at least 1500 bytes). The cache multiprogramming trace of Kaplan and Winder [17] and the (main memory) program address traces of Lewis and Shedler [18] and of Anacker and Wang ([16], [2]) give similar results. In the examples cited, different but similar page replacement algorithms were employed, including the working-set memory management policy [5] and the closely-related LRU ("least recently used") memory management policy ([19], [1]).

At first glance, some published data seem to contradict this insensitivity of miss ratio

Authors' present addresses R Fagin, IBM Research Laboratory, San Jose, CA 95193, M.C. Easton, IBM Thomas J. Watson Research Center, P O. Box 218, Yorktown Heights, NY 10598

[1] All requests in the trace were to the data base—there were no index references or program references.

to page size. A dramatic example is found in Chu and Opderbeck's paper [4], in which LRU miss ratio seems to depend very heavily on page size. Their miss ratio curves asymptotically approach a value which is simply the number of initial loading misses (which is the same as the number of pages in the program, if main memory is large enough to hold the entire program) divided by the length of the observed page reference string. Of course, in general, there are approximately twice as many initial loading misses when the page size is half as large, and so the asymptotic value of their miss ratio curve is approximately twice as big when the page size is half as large. By contrast, in this paper we distinguish between initial loading misses on the one hand, and the "transient-free," or "long-term," miss ratio on the other hand (cf. [9]). In the case of LRU, "transient-free" means that the miss ratio is measured starting at a time after main memory has filled; in the working-set case, "transient-free" means that the miss ratio is measured starting at a time greater than T, where T is the window size. In many cases, it is possible to analyze the performance of a storage hierarchy more accurately by considering the effects of initial loading misses and of the transient-free miss ratio separately. From here on in this paper, "miss ratio" refers to the transient-free, or long-term, miss ratio

The main result of this paper is a bound on the effect of page size on the expected working-set miss ratio, in the independent reference model (in which page i is referenced at time t with probability p_i, independent of past history). Specifically, let $MISS$ be the expected working-set miss ratio in this model, where the window size is chosen so that the expected working-set size is CAP pages. Now form blocks, by combining the B pages with the highest probabilities of reference into one block, the B pages with the next-highest probabilities of reference into a second block, and so on. Let $MISS^*$ be the expected working-set miss ratio when all data are moved in blocks, and when the window size is chosen so that the expected working-set size is again CAP pages, that is, $CAP/B = C$ blocks. It is proven that

$$|MISS - MISS^*| < 2/C + 33/C^2. \tag{1}$$

Thus, if C is sufficiently large, then $MISS^* \approx MISS$, that is, the expected miss ratios in the blocked and unblocked cases are approximately equal

In Section 2 we will show that (1) implies the approximate independence of miss ratio on page size in certain more realistic models of page references than the independent reference model. This tells us that in some cases, transient-free miss ratio is not a key factor in the selection of page size. Of course, the effect of initial loading misses, which can be considered separately, is an important factor.

2. *More Realistic Models*

In this section, we will show that our result about the insensitivity of miss ratio to page size in the independent reference model implies that this insensitivity holds in some more realistic models of page reference patterns. For convenience in exposition, we will deal in this section with LRU miss ratio, rather than with working-set miss ratio. Indeed, we will show later that our main result (1) can be interpreted as saying that LRU miss ratio is insensitive to page size in the independent reference model, if the capacity (size of main memory) in, say, bytes, is held fixed (and if, as before, pages are blocked together in order of their probabilities).

The independent reference model is, in general, inadequate. Various authors, including Lewis and Shedler [18], Denning, Savage, and Spirn [6], and Easton [7] have presented models of page reference patterns (or related quantities) in which page references are the result of two components, where the first is, roughly speaking, an "independent" or "random" component, and the second is a local component, such as a "locality of reference" or "sequential" component. We will first show that approximate independence of expected LRU miss ratio holds in one such model (Easton's), and then we will generalize the argument to cover other such models.

We briefly describe Easton's model. If there are n pages, then there are $n + 1$ parameters, $r, p_1, \cdots, p_n$, all between 0 and 1. Assume that page i was referenced at time t At time $t + 1$, a coin is flipped, which comes up heads with probability r, and tails with probability $1 - r$. If heads comes up, then page i is rereferenced, and we say that the reference to page i (at time $t + 1$) was generated during "rereference" or "sequential" mode. If tails comes up, page j is then referenced with probability p_j, for $1 \leq j \leq n$ (including the case $j = i$), and we say that the reference to page j was generated during "random" mode. Thus, if Q_{ij} is the probability that page j is referenced at time $t + 1$, given that page i was referenced at time t, then

$$Q_{ij} = \begin{cases} r + (1 - r)p_i, & j = i, \\ (1 - r)p_j, & j \neq i \end{cases}$$

Easton found [7] that with appropriate choice of parameters, his model gives a good fit to the LRU miss ratio curve of A.A.S., which we referred to in the Introduction. Intuitively, this model "works" because if the page size is large enough, then "locality" can be approximately captured by rereferences to the same page.

We will now show that in this model, LRU miss ratio is insensitive to page size. Let $S = R_1 \cdots R_m$ be a finite-length page reference string generated by this model. Thus, each R_i is the name of a page. Assume that the string S is long enough that the effect of initial loading misses is negligible. For each i, let R_i^* be the name of the block which contains page R_i. Then $S^* = R_1^* \cdots R_m^*$ is the corresponding block reference string. We will show that the LRU miss ratio (where pages are the unit of transfer) over page reference string S and with capacity CAP pages is approximately the same as the LRU miss ratio (where blocks are now the unit of transfer) over block reference string S^* and with capacity CAP/B blocks (i.e. CAP pages), when there are B pages per block.

Let $R_{i_1}, \cdots, R_{i_k}$ (where $i_1 < i_2 < \cdots < i_k$) be those page references which were generated after "tails" was flipped; thus, these are the page references which were generated during "random" mode. So $S_{\text{rand}} = R_{i_1} \cdots R_{i_k}$ is the substring of S which contains only the page references generated during random mode, and $S^*_{\text{rand}} = R_{i_1}^* \cdots R_{i_k}^*$ is the corresponding substring of S^*. Let z be the number of misses if we apply the LRU page replacement algorithm to string S_{rand}, with capacity CAP pages. Each reference which was generated during sequential mode is automatically a "hit"; hence, the number of LRU misses (with capacity CAP pages) over page reference string S is also z. Now S^*_{rand} can be looked at as a block reference string generated under the assumption of independent block references, where each block has reference probability equal to the sum of the reference probabilities of the pages which compose it. As we have said, our main result can be interpreted as saying that under the assumption of independent references, the LRU miss ratio is approximately the same in the blocked and unblocked cases. So z is also approximately the number of misses if we apply LRU block replacement to the block reference string S^*_{rand} with capacity CAP/B blocks. Again, there are exactly the same number of misses over the string S^* as over S^*_{rand}. So the miss ratios in the blocked and unblocked cases are approximately the same, as claimed.

This result can be generalized to some other models with an "independent" component and a "locality of reference" component. The argument we will now outline will, of course, have to be tailored to fit each particular model. Assume that the page size is large enough that almost all "locality" references are hits. Then the miss ratio is essentially determined by the independent component. So once again, the insensitivity of miss ratio to page size under the assumption of independent references implies this insensitivity under more realistic assumptions. We remark that for models other than the independent reference model, the difference in miss ratio between the blocked and unblocked cases will not necessarily be bounded by the right-hand side of inequality (1).

In our result about the independent reference model, we have assumed that pages

are blocked together in order of likelihood of reference. Of course, the real situation is much more complicated. However, we can justify this assumption on several grounds.

1. In the case of certain large-data base systems, such as A.A.S., groups of sequentially stored records tend to have similar access properties. So in forming blocks in the usual way of grouping together sequentially stored records, those pages which are blocked together have approximately the same probabilities of being referenced; that is, pages are blocked together approximately in order of probability of reference.

2. Of all possible ways of blocking pages together with B pages to a block, Yue and Wong [23] proved that in various storage applications and under various criteria of optimality, the blocking we have considered (in which pages are blocked together in order of probability of reference) is optimal. Hence, the use of this particular blocking is a natural assumption to make when one is discussing the performance of a storage hierarchy. We remark that the first author has found a counterexample to the conjecture that expected LRU miss ratio is minimized in the independent reference model by this blocking [10]; however, this blocking seems to be near-optimal in the independent reference model.

3. *Formal Statement of Main Result*

Let $\{p_1, \cdots, p_n\}$ be a probability distribution (that is, $\sum p_i = 1$ and each $p_i \geq 0$). Assume that at each discrete time t, page i is referenced with probability p_i, independent of past history. (This is the independent reference model.) The *expected working-set miss ratio (with window size T)* [5] is the probability that the page referenced at time t was not one of the pages referenced over the course of the previous T (not necessarily distinct) references. Under the independent, time-invariant assumption of the independent reference model, it is clear that this expected working-set miss ratio is independent of t, for $t > T$. Let CAP be the *expected working-set size*, that is, the expected number of distinct pages to appear over the course of T references. Define $MISS(CAP)$ to be the corresponding expected working-set miss ratio. Thus, $MISS(CAP)$ is the expected working-set miss ratio with window size T, where the expected working-set size with window size T is CAP pages. Later on, we will discuss the close relationship between $MISS(CAP)$ and the expected LRU miss ratio with capacity CAP pages.

We will now describe the blocked case. Let B (the "block size") be a positive integer which, for convenience, we assume divides n. Assume that $p_1 \geq p_2 \geq \cdots \geq p_n \geq 0$, and let

$$u_i = \sum_{j=1}^{B} p_{(i-1)B+j}, \qquad 1 \leq i \leq n/B.$$

Thus, $u_1 = p_1 + \cdots + p_B$, $u_2 = p_{B+1} + \cdots + p_{2B}$, etc. This corresponds to combining the B pages with the highest probabilities of reference into a block, the B pages with the next-highest probabilities of reference into a second block, and so on. The blocked case corresponds to the independent reference model with block probabilities $\{u_1, \cdots, u_{n/B}\}$. Define $MISS^*(CAP)$ to be the expected working-set miss ratio (over the probability distribution $\{u_1, \cdots, u_{n/B}\}$), when all data are moved in blocks, and when the window size is chosen so that the expected working-set size is CAP/B blocks (CAP/B blocks contain the same number of bytes as CAP pages, and this is the quantity we hold fixed in comparing the blocked and unblocked cases.)[2]

Let $C = CAP/B$, and write $MISS$ and $MISS^*$ for $MISS(CAP)$ and $MISS^*(CAP)$. The main result of this paper is

$$|MISS - MISS^*| < 2/C + 33/C^2. \tag{2}$$

[2] It may well happen that there is no integer T^* such that the expected working-set size with window size T^* is CAP/B blocks. If so, then we interpolate, as we will see

We can think of $2/C$ as the first-order error term, and $33/C^2$ as the second-order error term. We will actually prove a slightly stronger statement than (2).

Note that statement (2) is a distribution-free result: that is, the error terms do not depend on the values of the p_i (or even on n, the number of pages).

4. *Details About Main Result*

We begin this section by giving an expression [5] for $MISS(CAP)$, the expected working-set miss ratio when the expected working-set size is CAP pages. The expected working-set size, that is, the expected number of distinct pages which will be referenced over the course of T references, is $S(T) = \sum_{i=1}^{n} (1 - (1 - p_i)^T)$, because the probability that page i is referenced is $1 - (1 - p_i)^T$. The expected working-set miss ratio with window size T is $M(T) = \sum p_i (1 - p_i)^T$, because $p_i(1 - p_i)^T$ is the probability that page i is the next page referenced and that page i did not appear in the last T references. Thus, if $S(T) = CAP$, then $MISS(CAP)$ is by definition $M(T) = M(S^{-1}(CAP))$. Note that $M(S^{-1}(x))$ is well-defined for each real number x between 0 and n, even if the intermediate parameter $T = S^{-1}(x)$ is not an integer. By this procedure, which amounts to an interpolation, we can define $MISS(x) = M(S^{-1}(x))$ for each x with $0 \leq x < n$

Similarly, in the blocked case, we define $MISS^*(x) = M^*(S^{*-1}(x/B))$, for $0 \leq x < n$, where

$$S^*(T) = \sum_{i=1}^{n/B} (1 - (1 - u_i)^T), \qquad 0 \leq T < \infty,$$

$$M^*(T) = \sum_{i=1}^{n/B} u_i(1 - u_i)^T, \qquad 0 \leq T < \infty.$$

Thus, if the expected working-set size is CAP pages (i.e. CAP/B blocks), then $MISS^*(CAP)$ is the expected working-set miss ratio in the blocked case.

We will now briefly discuss the relationship between $MISS(CAP)$ on the one hand, and the expected LRU miss ratio $MR(CAP)$ with capacity CAP on the other hand. Denning and Schwartz [5] make the intuitive observation that $MISS(CAP) \approx MR(CAP)$. In various simulations of the independent reference model, we found that $MISS$ differs from MR by around .01, when the number n of pages is several hundred. Indeed, the first author has recently proven [11] that in a certain precise sense, $MISS$ converges asymptotically to MR as the number n of pages gets large, in the independent reference model. So (2) implies that in the independent reference model, the expected LRU miss ratio is approximately independent of page size, if the size of main memory is held fixed and if pages are blocked together in order of their probabilities. (Of course, we are assuming that C is large enough that the right-hand side of inequality (2) is small, and that n is large enough that $MISS(CAP) \approx MR(CAP)$.)

5. *Proof of Main Result*

We will prove the following theorem.

THEOREM. *Let $MISS$ be the expected working-set miss ratio in the independent reference model, with expected working-set size CAP pages. Let $MISS^*$ be the expected working-set miss ratio after blocking, where B old pages form each new block, where pages are blocked together in order of their reference probabilities, and where the expected working-set size is again CAP pages (i.e. $C = CAP/B$ blocks). Assume that the original number of pages is divisible by B. Then $|MISS - MISS^*| < 2/C + 33/C^2$.*

PROOF. The theorem is trivial if $CAP < 2B$ or $CAP \geq n$; hence, we will assume that $2B \leq CAP < n$.

We will actually prove

$$-1.92(B - 1)/CAP - 6(B/CAP)^2 < MISS - MISS^* < 1.01(B - 1)/CAP + 33(B/CAP)^2 \quad (3)$$

Of course, (3) implies $|MISS - MISS^*| < 2B/CAP + 33(B/CAP)^2$, which is the result of the theorem.

There will be six main steps in the proof.

Step 0. $S(T) \leq T$, if $T \geq 1$

Step 1. $0 \leq \sum_{i=1}^{n} (1 - p_i)^T - B \sum_{i=1}^{n/B} (1 - u_i/B)^T \leq B - 1$, if $T \geq 0$.

Step 2. $0 \leq B \sum_{i=1}^{n/B} (1 - u_i/B)^T - B \sum_{i=1}^{n/B} (1 - u_i)^{T/B} < .184\ (B - 1) + 6B^2/T$, if $T \geq 2B$.

Step 3. $-.736(B - 1)/T < \sum_{i=1}^{n} p_i(1 - p_i)^T - \sum_{i=1}^{n/B} u_i(1 - u_i/B)^T < .736(B - 1)/T$, if $T \geq 0$.

Step 4. $0 \leq \sum_{i=1}^{n/B} u_i(1 - u_i/B)^T - \sum_{i=1}^{n/B} u_i(1 - u_i)^{T/B} < .271(B - 1)/T + 33B^2/T^2$, if $T \geq 2B$.

Step 5. $H(T)/H(t) \leq S(T)/S(t)$, if $T \geq t \geq 1$, where $H(x) = 1 - M(x)$.

Step 0 follows immediately from the development in [5], provided T is an integer (which will not always be the case for us—hence, we must prove it directly).

Step 5 says that if the expected working-set size is increased, then the proportional increase in expected working-set hit ratio is bounded by the proportional increase in expected working-set size. Thus, if the expected working-set size is doubled, then the expected hit ratio is at most doubled. We will now show that these six steps imply statement (3).

Instead of using the functions S^* and M^* of Section 2, it will be convenient to define closely related functions S_B and M_B, as follows.

$$S_B(T) = B \sum_{i=1}^{n/B} (1 - (1 - u_i)^{T/B}), \qquad M_B(T) = \sum_{i=1}^{n/B} u_i(1 - u_i)^{T/B}.$$

It is easy to see that $M_B(S_B^{-1}(CAP)) = M^*(S^{*-1}(CAP/B)) = MISS^*(CAP)$, the expected working-set miss ratio under blocking.

Two other functions we will find convenient to use are given by

$$H(T) = 1 - M(T), \qquad H_B(T) = 1 - M_B(T).$$

It is easy to see that S and S_B are each monotone, and each map onto the half-closed, half-open interval $[0, n)$. So we can find T_1 and T_B such that $S(T_1) = CAP = S_B(T_B)$. Now $T_1 \geq 1$, since $S(1) = 1 < CAP = S(T_1)$, and since S is monotone increasing. So we can apply step 0, to obtain $T_1 \geq S(T_1) = CAP \geq 2B$. Hence

$$T_1 \geq 2B. \tag{4}$$

Now $S(T) = n - \sum_{i=1}^{n} (1 - p_i)^T$, and $S_B(T) = n - B \sum_{i=1}^{n/B} (1 - u_i)^{T/B}$. So if we add together the inequalities of steps 1 and 2, with T_1 substituted for T, we find that

$$0 \leq S_B(T_1) - S(T_1) < Q, \tag{5}$$

where

$$Q = 1.184\ (B - 1) + 6\ (B^2/T_1). \tag{6}$$

Since $S(T_1) = CAP$, statement (5) says

$$0 \leq S_B(T_1) - CAP < Q. \tag{7}$$

So $S_B(T_1) \geq CAP = S_B(T_B)$. Since S_B is monotone increasing,

$$T_1 \geq T_B. \tag{8}$$

The functions H_B and S_B/B have the same form as H and S, with u_i substituted for p_i, with n/B substituted for n, and with T/B substituted for T. So it follows from step 5 that if $(T/B) \geq (t/B) \geq 1$, that is, if $T \geq t \geq B$, then

$$H_B(T)/H_B(t) \leq (S_B(T)/B)/(S_B(t)/B) = S_B(T)/S_B(t). \tag{9}$$

Now $T_1 \geq T_B \geq B$: first, $T_1 \geq T_B$ by (8), and $S_B(T_B) = CAP > B = S_B(B)$, so since S_B is monotone increasing, this implies that $T_B > B$. Therefore, we can substitute T_1 for T and T_B for t in (9); then

$$H_B(T_1)/H_B(T_B) \leq S_B(T_1)/S_B(T_B) = S_B(T_1)/CAP.$$

So

$$\begin{aligned} H_B(T_1) &\leq H_B(T_B)(S_B(T_1)/CAP) \\ &= H_B(T_B)(1 + ((S_B(T_1) - CAP)/CAP)) \\ &\leq H_B(T_B)(1 + (Q/CAP)) \quad \text{by (7)} \\ &\leq H_B(T_B) + (Q/CAP), \quad \text{since } H_B(T_B) \leq 1 \end{aligned}$$

We have just shown that

$$H_B(T_1) \leq H_B(T_B) + (Q/CAP). \tag{10}$$

It is easy to see that H_B is monotone increasing. So, from (8),

$$H_B(T_1) \geq H_B(T_B). \tag{11}$$

From (10) and (11), it immediately follows that

$$0 \leq H_B(T_1) - H_B(T_B) \leq Q/CAP,$$

and hence

$$-Q/CAP \leq M_B(T_1) - M_B(T_B) \leq 0. \tag{12}$$

Adding together the inequalities of steps 3 and 4, with T_1 substituted for T (which is all right by (4)), we find that

$$-.736\ (B - 1)/T_1 < M(T_1) - M_B(T_1) < 1.01\ (B - 1)/T_1 + 33\ (B^2/T_1^2). \tag{13}$$

Since $T_1 \geq CAP$ by step 0, statement (13) implies

$$\begin{aligned} -.736\ (B - 1)/CAP < M(T_1) - M_B(T_1) \\ < 1.01\ (B - 1)/CAP + 33\ (B/CAP)^2. \end{aligned} \tag{14}$$

Also, $T_1 \geq CAP$ implies (from statement (6)) that

$$Q \leq 1.184\ (B - 1) + 6\ (B^2/CAP). \tag{15}$$

If we add together the inequalities in statements (12) and (14), and substitute for Q the right-hand side of (15), we get

$$\begin{aligned} -1.92\ (B - 1)/CAP - 6\ (B/CAP)^2 < M(T_1) - M_B(T_B) \\ < 1.01\ (B - 1)/CAP + 33\ (B/CAP)^2. \end{aligned}$$

Since as we observed, $M(T_1) = MISS(CAP)$ and $M_B(T_B) = MISS^*(CAP)$, this gives us statement (3), as desired.

It remains to prove steps 0–5.

6. *Preliminaries*

One of our basic techniques will be the use of Schur functions ([21]; see also [20]). Assume for convenience throughout that all functions considered are infinitely differentiable

Definition. Assume that $\alpha = (\alpha_1, \cdots, \alpha_m)$, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m$, and that $\beta = (\beta_1, \cdots, \beta_m)$, where $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m$. We say $\alpha \succ \beta$ if $\sum_{i=1}^{k} \alpha_i \geq \sum_{i=1}^{k} \beta_i$, $1 \leq k \leq m$, and $\sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} \beta_i$.

Definition. A real-valued function f of m real variables $x_1, \cdots, x_m$ is a *Schur function* if for each pair $i \neq j$, $(x_i - x_j)(\partial/\partial x_i)f - (\partial/\partial x_j)f \geq 0$.

THEOREM (Schur). *Let f be defined for $x_1 \geq \cdots \geq x_m$. Then $f(\alpha_1, \cdots, \alpha_m) \geq f(\beta_1, \cdots, \beta_m)$ whenever $\alpha \succ \beta$, iff f is a Schur function.*

Special Case 1 [13, p. 89]. Let ϕ be a real convex function of one real variable, that is,

$$\phi((x + y)/2) \leq \tfrac{1}{2}(\phi(x) + \phi(y)). \tag{16}$$

Then the function $(x_1, \cdots, x_m) \to \sum_{i=1}^{m} \phi(x_i)$ is a Schur function. Hence, if $\alpha \succ \beta$, then $\sum_{i=1}^{m} \phi(\alpha_i) \geq \sum_{i=1}^{m} \phi(\beta_i)$.

Special Case 2. Let ϕ be continuous and concave, that is, the inequality sign in (16) is reversed. Then $-\phi$ is convex. Hence, if $\alpha \succ \beta$, then $\sum_{i=1}^{m} \phi(\alpha_i) \leq \sum_{i=1}^{m} \phi(\beta_i)$.

An infinite sum $\sum_{i=1}^{\infty} a_i$ of real numbers a_i is *alternating* if

1. $a_i \geq 0$ *iff* $a_{i+1} \leq 0$, $1 \leq i < \infty$.
2. $|a_i| \geq |a_{i+1}|$, $1 \leq i < \infty$.
3. $a_i \to 0$ as $i \to \infty$.

An alternating sum is always convergent, and its value lies between any two successive partial sums (*Leibnitz's test* [15, p. 68]).

7. *Proof of Step* 0

We will prove that $S(T) \leq T$, if $T \geq 1$.

Let ϕ be the function $x \to 1 - (1 - x)^T$, with domain the closed interval [0, 1]. It is easy to see that ϕ is concave. Clearly $(p_1, \cdots, p_n) \succ (1/n, \cdots, 1/n)$. So by special case 2 of Section 6,

$$S(T) = \sum_{i=1}^{n} \phi(p_i) \leq \sum_{i=1}^{n} \phi(1/n) = n\phi(1/n).$$

Hence, we need only show that $n\phi(1/n) \leq T$. This is equivalent to

$$(1 - 1/n)^T \geq 1 - T/n. \tag{17}$$

If $T > n$, then the right-hand side of (17) is negative, and (17) follows immediately. If $T \leq n$, then the binomial expansion of $(1 - 1/n)^T$ is an alternating sum, and (17) again follows.

8. *Proof of Step* 1

We will show that

$$0 \leq (1/B) \sum_{i=1}^{n} (1 - p_i)^T - \sum_{i=1}^{n/B} (1 - u_i/B)^T \leq (B - 1)/B, \tag{18}$$

if $T \geq 0$.

Let ϕ be the function $x \to (1 - x)^T$, with domain the closed interval [0, 1]. It is easy to see that ϕ is convex and monotone decreasing with range [0, 1]. Since ϕ is convex, it follows [13, p. 72] that

$$\phi((x_1 + \cdots + x_B)/B) \leq (\phi(x_1) + \cdots + \phi(x_B))/B. \tag{19}$$

If we substitute $x_1 = p_{(i-1)B+1}$, $x_2 = p_{(i-1)B+2}$, $\cdots$, $x_B = p_{iB}$ into (19), for $1 \leq i \leq n/B$, we find that

$$(1 - u_i/B)^T \leq (1/B)((1 - p_{(i-1)B+1})^T + \cdots + (1 - p_{iB})^T). \tag{20}$$

Adding together the inequalities (20), for $1 \leq i \leq n/B$, we obtain

$$\sum_{i=1}^{n/B} (1 - u_i/B)^T \leq (1/B) \sum_{i=1}^{n} (1 - p_i)^T,$$

which proves the first inequality in (18).

If $1 \geq x_1 \geq \cdots \geq x_B \geq 0$, then since ϕ is monotone decreasing,

$$\frac{\phi(x_1) + \cdots + \phi(x_B)}{B} - \phi\left(\frac{x_1 + \cdots + x_B}{B}\right) \leq \frac{(B - 1)\phi(x_B) + \phi(x_1)}{B} - \phi(x_1) = \frac{B - 1}{B}(\phi(x_B) - \phi(x_1)). \tag{21}$$

Substituting $x_1 = p_{(i-1)B+1}$, $x_2 = p_{(i-1)B+2}$, $\cdots$, $x_B = p_{iB}$ into (21), for $1 \leq i \leq n/B$, we obtain

$$(1/B)(\phi(p_{(i-1)B+1}) + \cdots + \phi(p_{iB})) - \phi(u_i/B) \leq ((B-1)/B)(\phi(p_{iB}) - \phi(p_{(i-1)B+1})). \quad (22)$$

Now $\phi(p_{(i-1)B+1}) \geq \phi(p_{(i-1)B})$, so we obtain from (22)

$$(1/B)(\phi(p_{(i-1)B+1}) + \cdots + \phi(p_{iB})) - \phi(u_i/B) \leq ((B-1)/B)(\phi(p_{iB}) - \phi(p_{(i-1)B})). \quad (23)$$

Adding together inequality (22) for $i = 1$ to inequalities (23) for $1 < i \leq n/B$, the right-hand side telescopes to give

$$(1/B)\sum_{i=1}^{n}\phi(p_i) - \sum_{i=1}^{n/B}\phi(u_i/B) \leq ((B-1)/B)(\phi(p_n) - \phi(p_1)) \leq (B-1)/B. \quad (24)$$

This is the right-hand inequality of (18).

9. *Proof of Step 2*

We will show that

$$0 \leq B\sum_{i=1}^{n/B}(1 - u_i/B)^T - B\sum_{i=1}^{n/B}(1-u_i)^{T/B} < .184(B-1) + 6(B^2/T), \quad (25)$$

if $T \geq 2B$.

We will first demonstrate the first inequality. We need only show that for each i,

$$(1 - u_i/B)^T \geq (1-u_i)^{T/B}. \quad (26)$$

This is equivalent to showing that

$$(1 - u_i/B)^B \geq (1-u_i). \quad (27)$$

It is straightforward to check that the binomial expansion of the left-hand side of (27) is an alternating sum. So (27) follows

We will now prove the second inequality of (25). For each u, $0 \leq u \leq 1$, and each nonnegative number θ, define

$$a_\theta(u) = (1 - u/B)^{B\theta}, \quad b_\theta(u) = (1-u)^\theta, \quad \epsilon_\theta(u) = a_\theta - b_\theta. \quad (28)$$

If $\frac{1}{2} < u \leq 1$, then $\epsilon_\theta(u) \leq a_\theta(u) \leq (1 - 1/2B)^{B\theta}$.

In particular, if $\theta = T/B$, then

$$\epsilon_{T/B}(u) \leq (1 - 1/2B)^T. \quad (29)$$

Assume from here on that $0 \leq u \leq \frac{1}{2}$. We will write ϵ_θ for $\epsilon_\theta(u)$, etc Let x and y be nonnegative real numbers. Then inequality (26), with Bx substituted for T and u substituted for u_i, gives $a_x \geq b_x$. Hence

$$a_x + b_x \leq 2a_x. \quad (30)$$

Multiplying together inequality (30) and the equality $a_y - b_y = \epsilon_y$, we obtain

$$a_xa_y + a_yb_x - a_xb_y - b_xb_y \leq 2a_x\epsilon_y. \quad (31)$$

Clearly,

$$a_xa_y = a_{x+y}, \qquad b_xb_y = b_{x+y}. \quad (32)$$

Substituting into (31) using (32), and replacing $a_{x+y} - b_{x+y}$ by ϵ_{x+y} and rearranging terms, we obtain

$$\epsilon_{x+y} \leq 2a_x\epsilon_y + a_xb_y - a_yb_x. \quad (33)$$

Substituting $a_x - \epsilon_x$ for b_x and $a_y - \epsilon_y$ for b_y in (33), we find

$$\epsilon_{x+y} \leq \epsilon_x a_y + \epsilon_y a_x . \tag{34}$$

We now claim that if j is a positive integer, then

$$\epsilon_j \leq j\epsilon_1 a_1^{j-1} \tag{35}$$

This is obvious if $j = 1$. Assume inductively that it is true for $j = N$ Then from (34),

$$\begin{aligned} \epsilon_{N+1} &\leq \epsilon_N a_1 + \epsilon_1 a_N \leq N\epsilon_1 a_1^N + \epsilon_1 a_N \quad \text{by inductive assumption} \\ &= N\epsilon_1 a_1^N + \epsilon_1 a_1^N = (N+1)\epsilon_1 a_1^N. \end{aligned}$$

Hence (35) holds, by induction, for every positive integer j.

Write $T/B = j + r$, j a positive integer and $0 \leq r < 1$. Then

$$\begin{aligned} \epsilon_{T/B} &= \epsilon_{j+r} \\ &\leq \epsilon_j a_r + \epsilon_r a_j \qquad \text{by (34)} \\ &\leq j\epsilon_1 a_1^{j-1} a_r + \epsilon_r a_j \qquad \text{by (35)} \\ &= j\epsilon_1 a_1^{j+r-1} + \epsilon_r a_1^j \qquad \text{since } a_r = a_1^r \text{ and } a_j = a_1^j \\ &\leq j\epsilon_1 a_1^{j+r-1} + \epsilon_r a_1^{j+r-1} \qquad \text{since } a_1 < 1 \text{ and } r < 1 \\ &\leq (T/B)\epsilon_1 a_1^{(T/B)-1} + \epsilon_r a_1^{(T/B)-1} \\ &= (T/B)\epsilon_1 (1 - u/B)^{T-B} + \epsilon_r (1 - u/B)^{T-B} \end{aligned} \tag{36}$$

How big is ϵ_r? As we observed, the binomial expansion of $(1 - u/B)^B$ is an alternating sum. Hence

$$(1 - u/B)^B \leq 1 - u + (B-1)u^2/2B. \tag{37}$$

So

$$\begin{aligned} \epsilon_r &= a_r - b_r = (1 - u/B)^{Br} - (1-u)^r \\ &\leq (1 - u + (B-1)u^2/2B)^r - (1-u)^r \quad \text{by (37)} \\ &= (1-u)^r[(1 + (B-1)u^2/2B(1-u))^r - 1]. \end{aligned} \tag{38}$$

Let $z = (B-1)u^2/2B(1-u)$.

Since $0 \leq u \leq \frac{1}{2}$, it is easy to see that $0 \leq z \leq 1$. Now the binomial expansion of $(1+z)^r - 1$ is an alternating sum if $0 \leq z < 1$ and $0 \leq r < 1$ (that is, there is an alternating sum if we consider only the second, third, $\cdots$ terms of the binomial expansion of $(1+z)^r$). Putting this together with (38), we obtain

$$\begin{aligned} \epsilon_r &\leq (1-u)^r[r(B-1)u^2/2B(1-u)] = r(B-1)u^2/2B(1-u)^{1-r} \\ &\leq r(B-1)u^2/2^r B \quad \text{since} \quad (1-u) \geq \tfrac{1}{2} \\ &\leq .54\ ((B-1)u^2/B), \end{aligned}$$

since we find by elementary calculus that the maximum of $r/2^r$, $0 \leq r \leq 1$, is $1/(e \log 2) < .54$. So

$$\epsilon_r \leq .54(B-1)u^2/B \leq .54\, u^2. \tag{39}$$

How big is ϵ_1 ? From (37), we find immediately that

$$\epsilon_1 \leq (B-1)u^2/2B. \tag{40}$$

If we substitute into the last line of (36) using (39) and (40), we obtain

$$\epsilon_{T/B} \leq u^2(1 - u/B)^{T-B}((T(B-1)/2B^2) + .54). \tag{41}$$

We are interested in obtaining an upper bound for $B\sum_{i=1}^{n/B} \epsilon_{T/B}(u_i)$. At most one u_i, namely u_1, can be greater than $\frac{1}{2}$. So from (29) and (41),

$$\begin{aligned} B\sum_{i=1}^{n/B} \epsilon_{T/B}(u_i) &< B(1 - 1/2B)^T \\ &\quad + ((T(B-1)/2B) + .54B)\sum_{i=1}^{n/B} u_i^2(1 - u_i/B)^{T-B}. \end{aligned} \tag{42}$$

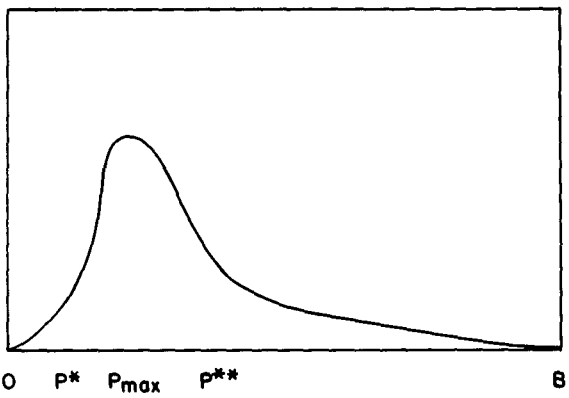

FIG. 1

How big can $\sum_{i=1}^{B/n} u_i^2(1 - u_i/B)^{T-B}$ be? Let ϕ be the function $x \to x^2(1 - x/B)^{T-B}$, with domain $[0, B]$. By using elementary calculus, we find that ϕ has its maximum at $P_{max} = 2B/(T + 2 - B)$, and two points of inflection; the first such point of inflection, P^*, lies strictly between 0 and P_{max}, and the second point of inflection, P^{**}, is bigger than P_{max}. The second derivative of ϕ is positive between 0 and P^*; negative, between P^* and P^{**}; and positive, between P^{**} and B. So the shape of ϕ is roughly as in Figure 1.

Let $VEC_1 = \{(x_1, \cdots, x_m) : m$ a positive integer, $\sum_{i=1}^m x_i \leq 1$, and $1 \geq x_1 \geq x_2 \geq \cdots \geq x_m \geq 0\}$. Note that the length of a tuple in VEC_1 is arbitrary (but finite). It is not important that the x_i's are in descending order, but it will make the exposition simpler. Say we can find an upper bound M for $\{\sum \phi(x_i) : \mathbf{x} \in VEC_1\}$, where $\mathbf{x}$ is an abbreviation for $(x_1, \cdots, x_m)$. Then $\sum_{i=1}^{n/B} u_i^2(1 - u_i/B)^{T-B} \leq M$.

Let $VEC_2 = \{\mathbf{x} : \sum x_i \leq 1$, and $P_{max} \geq x_1 \geq x_2 \geq \cdots \geq x_m \geq 0\}$. Then $\sup\{\sum \phi(x_i) : \mathbf{x} \in VEC_1\} = \sup\{\sum \phi(x_i) : \mathbf{x} \in VEC_2\}$, since if $\mathbf{x} \in VEC_1$ and $y_i = \min(x_i, P_{max})$ for each i, then $\mathbf{y} \in VEC_2$.

Let $\mathbf{v}$ be a tuple in VEC_2; then $\mathbf{v}$ can be written as the concatenation $\widehat{\mathbf{yz}}$ of a tuple $\mathbf{y}$ (with all entries between P^* and P_{max}) and a tuple $\mathbf{z}$ (with all entries between 0 and P^*); if any entry of $\mathbf{x}$ is exactly P^*, we place that entry in $\mathbf{z}$. Assume that $\sum z_i = a$. Let l be the unique nonnegative integer and δ the unique real number, $0 \leq \delta < P^*$, such that $a = lP^* + \delta$. Let $\boldsymbol{\alpha}$ be the vector $(P^*, \cdots, P^*, \delta, 0, \cdots, 0)$ with length the same as that of $\mathbf{z}$, and with l occurrences of P^*, one occurrence of δ, and the rest 0's. Clearly, $\boldsymbol{\alpha} > \mathbf{z}$. Now $\phi''(x) \geq 0$ for each x in $[0, P^*]$; hence [13, p. 76] ϕ is convex in $[0, P^*]$. Applying special case 1 of Section 5, we find that $\sum_{i=1}^k \phi(\alpha_i) \geq \sum_{i=1}^k \phi(z_i)$. We have shown that if $\mathbf{v} = \widehat{\mathbf{yz}}$ is an arbitrary tuple in VEC_1, then there is a tuple $\mathbf{w} = \widehat{\mathbf{y}}(P^*, \cdots, P^*, \delta)$ such that $\sum \phi(v_i) \leq \sum \phi(w_i)$. In other words: let

$$VEC_3 = \{\mathbf{x} : \sum x_i \leq 1, P_{max} \geq x_1 \geq x_2 \geq \cdots \geq x_{m-1} \geq P^*, P_{max} \geq x_m \geq 0,$$

where m is the length of $\mathbf{x}\}$. Then $\sup\{\sum \phi(x_i) : \mathbf{x} \in VEC_3\} = \sup\{\sum \phi(x_i) : \mathbf{x} \in VEC_2\}$.

Now let $\mathbf{v}$ be an arbitrary tuple in VEC_3. Write $\mathbf{v} = \widehat{\mathbf{yz}}$, where each entry of $\mathbf{y}$ lies between P^* and P_{max}, and where $\mathbf{z}$ is a tuple of length 1, whose entry lies between 0 and P_{max}. Assume that $\mathbf{y}$ is of length k, and that $\sum y_i = a$. Let $\boldsymbol{\alpha}$ be the tuple $(a/k, \cdots, a/k)$ of length k. Then $\mathbf{y} > \boldsymbol{\alpha}$. Now $\phi''(x) \leq 0$ for each x in $[P^*, P_{max}]$; hence, ϕ is concave in $[P^*, P_{max}]$. Applying special case 2 of Section 5, we find that $\sum \phi(y_i) \leq \sum \phi(\alpha_i)$. In other words: let $VEC_4 = \{\mathbf{x} : \sum x_i \leq 1, x_1 = \cdots = x_{m-1}$, where m is the length of $\mathbf{x}\}$. Then $\sup\{\sum \phi(x_i) : \mathbf{x} \in VEC_3\} = \sup\{\sum \phi(x_i) : \mathbf{x} \in VEC_4\}$.

Let $\mathbf{v} = \{\beta, \beta, \cdots, \beta, \delta\}$ be an arbitrary tuple in VEC_4. Then

$$\sum \phi(v_i) \leq (1/\beta)\lfloor\phi(\beta)\rfloor + \phi(\delta) \leq (\phi(\beta)/\beta) + \phi(\delta)$$
$$\leq (\max_{0 \leq x \leq 1} \phi(x)/x) + (\max_{0 \leq x \leq 1} \phi(x)). \quad (43)$$

Now $\phi(x)/x = x(1 - x/B)^{T-B}$ Let k be arbitrary for the moment, and let ψ be the function $x \to x^k(1 - x/B)^{T-B}$. We are interested in finding $\max_{0 \le x \le 1} \psi$ for $k = 1$ and $k = 2$ (and, in proving step 4, we will be interested in the case $k = 3$). By elementary calculus, we find that ψ has its maximum at $kB/(T + k - B)$, with maximum value

$$(kB/(T + k - B))^k (1 - k/(T + k - B))^{T-B}. \tag{44}$$

We will now estimate $(1 - k/(T + k - B))^{T-B}$.

It is well known that $(1 - 1/x)^x \to 1/e$ as $x \to \infty$ Also, the function $x \to (1 - 1/x)^x$ is a monotone, strictly increasing function of x for $x > 1$, for, the derivative of this function, evaluated at $x > 1$, is easily found to be

$$\begin{aligned}
&\left(1 - \frac{1}{x}\right)^{x-1}\left[\left(1 - \frac{1}{x}\right)\log\left(1 - \frac{1}{x}\right) + \frac{1}{x}\right] \\
&\quad = \left(1 - \frac{1}{x}\right)^{x-1}\left[\left(1 - \frac{1}{x}\right)\left(-\frac{1}{x} - \frac{1}{2x^2} - \frac{1}{3x^3} - \frac{1}{4x^4} - \cdots\right) + \frac{1}{x}\right] \\
&\qquad\qquad \text{by using the Taylor expansion of } \log(1 - 1/x) \\
&\quad = \left(1 - \frac{1}{x}\right)^{x-1}\left[\left(1 - \frac{1}{2}\right)\frac{1}{x^2} + \left(\frac{1}{2} - \frac{1}{3}\right)\left(\frac{1}{x^3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right)\frac{1}{x^4} + \cdots\right] \\
&\quad > 0
\end{aligned}$$

So

$$\frac{1}{e} > \left(1 - \frac{1}{x+1}\right)^{x+1} = \left(1 - \frac{1}{x+1}\right)^{x}\left(1 - \frac{1}{x+1}\right) = \left(1 - \frac{1}{x+1}\right)^{x}\frac{x}{x+1}. \tag{45}$$

If we multiply (45) through by $(x + 1)/x$, we find that

$$\left(1 - \frac{1}{x+1}\right)^{x} < \frac{1}{e}\frac{x+1}{x}. \tag{46}$$

So

$$\begin{aligned}
\left(1 - \frac{k}{T + k - B}\right)^{T-B} &= \left(1 - \frac{1}{x+1}\right)^{xk}, \quad \text{where } x = (T - B)/k \\
&< \left(\frac{1}{e} \cdot \frac{x+1}{x}\right)^k \quad \text{by (46)} \\
&= \left(\frac{T + k - B}{e(T - B)}\right)^k.
\end{aligned} \tag{47}$$

So the maximum value of ψ, which we found to be expression (44), is bounded by (from(47))

$$\left(\frac{kB}{T + k - B}\right)^k \left(\frac{T + k - B}{e(T - B)}\right)^k = \left(\frac{kB}{e(T - B)}\right)^k. \tag{48}$$

So from (43),

$$\sum \phi(v_i) \le \frac{B}{e(T - B)} + \frac{4B^2}{e^2(T - B)^2}.$$

Tracing back what we have demonstrated, this means that

$$\sum_{i=1}^{n/B} u_i^2 \left(1 - \frac{u_i}{B}\right)^{T-B} \le \frac{B}{e(T - B)} + \frac{4B^2}{e^2(T - B)^2}. \tag{49}$$

We will now estimate the subexpression $(1 - 1/2B)^T$ of (42). It will be convenient for work later on to find a more general estimate than we need now. We will show that if $s \ge 1$ and $t > 0$, then

$$(1 - 1/s)^t < m! s^m/t^m \quad \text{for each integer } m \ge 0. \tag{50}$$

First,

$$(1 - 1/s)^t = (1 - 1/s)^{s(t/s)} < e^{-t/s}, \tag{51}$$

since as we showed, $(1 - 1/x)^x \nearrow 1/e$. Now if $z > 0$, then $e^z > z^m/m!$ for each integer $m \geq 0$, since $z^m/m!$ is one term of the Taylor expansion of e^z. Hence $e^{-z} < m!/z^m$. Applying this to the right-hand term of (51), $(1 - 1/s)^t < e^{-t/s} < m!\, s^m/t^m$, as desired.

Hence

$$(1 - 1/2B)^T < 2B/T, \tag{52}$$

where we let $m = 1$ in (50).

Substituting into (42) using (49) and (52), we obtain

$$\begin{aligned} B \sum_{i=1}^{n/B} \epsilon_{T/B}(u_i) &< \frac{2B^2}{T} + \left(\frac{T(B-1)}{2B} + .54B\right)\left(\frac{B}{e(T-B)} + \frac{4B^2}{e^2(T-B)^2}\right) \\ &= \frac{2B^2}{T} + \frac{(B-1)T}{2e(T-B)} + \frac{.54B^2}{e(T-B)} + \frac{2BT(B-1)}{e^2(T-B)^2} + \frac{2.16B^3}{e^2(T-B)^2} \\ &\leq \frac{2B^2}{T} + \frac{B-1}{2e}\,\frac{T}{T-B} + \frac{.54B}{e}\,\frac{B}{T-B} \\ &\qquad + \frac{2B}{e^2}\,\frac{T}{T-B}\,\frac{B}{T-B} + \frac{2.16B}{e^2}\,\frac{B}{T-B}\,\frac{B}{T-B}. \end{aligned} \tag{53}$$

We will find simple upper bounds for $T/(T - B)$ and $B/(T - B)$, given that $T \geq 2B$.

$$\frac{B}{T-B} = \frac{B}{T}\left(1 + \frac{B}{T-B}\right) \leq \frac{2B}{T} \quad \text{since } B/(T-B) \leq 1, \tag{54}$$

$$\frac{T}{T-B} = 1 + \frac{B}{T-B} \leq 1 + \frac{2B}{T} \quad \text{by (54)}. \tag{55}$$

Substituting into (53) using (54) and (55), we obtain

$$\begin{aligned} B \sum_{i=1}^{n/B} \epsilon_{T/B}(u_i) &< \frac{2B^2}{T} + \frac{B-1}{2e}\left(1 + \frac{2B}{T}\right) + \frac{.54B}{e}\,\frac{2B}{T} \\ &\qquad + \frac{2B}{e^2}\left(1 + \frac{2B}{T}\right)\left(\frac{2B}{T}\right) + \frac{2.16B}{e^2}\left(\frac{2B}{T}\right)\left(\frac{2B}{T}\right) \\ &< \frac{B-1}{2e} + \frac{B^2}{T}\left(2 + \frac{1}{e} + \frac{1.08}{e} + \frac{4}{e^2}\right) + \frac{B^3}{T^2}\left(\frac{8}{e^2} + \frac{8.64}{e^2}\right), \end{aligned} \tag{56}$$

where we have expanded out and replaced the term $\frac{B-1}{e}\frac{B}{T}$ by $\frac{B^2}{eT}$.

If we replace B^3/T^2 on the right-hand side of (56) by B^2/T, and numerically evaluate, we find

$$B \sum_{i=1}^{n/B} \epsilon_{T/B}(u_i) < .184\,(B-1) + 5.56\,(B^2/T),$$

which implies the second inequality of (25).

10. *Proof of Step 3*

We will show that

$$\left| \sum_{i=1}^{n} p_i(1-p_i)^T - \sum_{i=1}^{n/B} u_i(1-u_i/B)^T \right| < .736(B-1)/T, \tag{57}$$

if $T \geq 0$.

Let ψ be the function $x \to x(1-x)^T$, with domain $[0, 1]$. We find by elementary calculus that ψ has its maximum at $\mu = 1/(T+1)$, with maximum value

$$\begin{aligned} \text{MAX} &= \frac{1}{T+1}\left(1 - \frac{1}{T+1}\right)^T \\ &< \frac{1}{eT} \quad \text{by (46)}. \end{aligned} \tag{58}$$

Further, ψ is monotone increasing between 0 and μ, and monotone decreasing between μ and 1.

If $1 \geq x_1 \geq \cdots \geq x_B \geq \mu$, then since ψ is monotone decreasing between μ and 1, we find as in (21) of the proof of step 1 that

$$((\psi(x_1) + \cdots + \psi(x_B))/B) - \psi((x_1 + \cdots + x_B)/B) \leq ((B-1)/B)(\psi(x_B) - \psi(x_1)). \tag{59}$$

And,

$$\begin{aligned}(\psi(x_1 + \cdots + x_B)/B) - (\psi(x_1) + \cdots + \psi(x_B))/B &\leq \psi(x_B) - ((B-1)\psi(x_1) + \psi(x_B))/B \\ &= ((B-1)/B)(\psi(x_B) - \psi(x_1)).\end{aligned} \tag{60}$$

Putting together (59) and (60), we obtain

$$|(\psi(x_1) + \cdots + \psi(x_B))/B - \psi((x_1 + \cdots + x_B)/B)| \leq ((B-1)/B)(\psi(x_B) - \psi(x_1)). \tag{61}$$

Let k be the maximal integer such that $p_{kB} \geq \mu$. Then $1 \geq p_1 \geq p_2 \geq \cdots \geq p_{kB} \geq \mu$. By using (61) in an analogous way to our use of (21) in the proof of step 1, we obtain, as in (24) of step 1, that

$$\begin{aligned}(1/B)\left|\sum_{i=1}^{kB} \psi(p_i) - B\sum_{i=1}^{k} \psi(u_i/B)\right| &\leq ((B-1)/B)(\psi(p_{kB}) - \psi(p_i)) \\ &\leq ((B-1)/B)\psi(p_{kB}).\end{aligned} \tag{62}$$

We know that $p_{(k+1)B} < \mu$. There are now two cases to consider.

Case 1. $p_{kB+1} \leq \mu$. Assume that $\mu \geq x_1 \geq x_2 \geq \cdots \geq x_B \geq 0$. Since ψ is monotone increasing between 0 and μ, we find, by a similar argument to that used to prove (61) and (62), that

$$|(\psi(x_1) + \cdots + \psi(x_B))/B - \psi((x_1 + \cdots + x_B)/B)| \leq ((B-1)/B)(\psi(x_1) - \psi(x_B)). \tag{63}$$

$$(1/B)\left|\sum_{i=kB+1}^{n} \psi(p_i) - B\sum_{i=1}^{k} \psi(u_i/B)\right| \leq ((B-1)/B)\psi(p_{kB+1}). \tag{64}$$

Hence, from (62) and (64),

$$\begin{aligned}(1/B)\left|\sum_{i=1}^{n} \psi(p_i) - B\sum_{i=1}^{n/B} \psi(u_i/B)\right| &\leq ((B-1)/B)(\psi(p_{kB}) + \psi(p_{kB+1})) \\ &\leq (2(B-1)/B)\text{MAX} \\ &\leq 2(B-1)/eBT \quad \text{by (58).}\end{aligned}$$

Since $2/e < .736$, this gives us (57).

Case 2. $p_{kB+1} > \mu$. We will show that

$$|(\psi(p_{kB+1}) + \cdots + \psi(p_{(k+1)B}))/B - \psi(u_{k+1}/B)| \leq ((B-1)/B)(\text{MAX} - \psi(p_{kB+1})) + ((B-1)/B)(\text{MAX} - \psi(p_{(k+1)B})). \tag{65}$$

We will first show that this is sufficient to prove (57).

As in the proof of (64) of case 1,

$$(1/B)\left|\sum_{i=p_{(k+1)B+1}}^{n} \psi(p_i) - B\sum_{i=k+2}^{n/B} \psi(u_i/B)\right| \leq ((B-1)/B)\psi(p_{(k+1)B+1}). \tag{66}$$

Since $p_{kB+1} > \mu$, and since ψ is monotone decreasing between u and 1, it follows that $\psi(p_{kB}) \leq \psi(p_{kB+1})$. So from (62),

$$(1/B)\left|\sum_{i=1}^{kB} \psi(p_i) - B\sum_{i=1}^{k} \psi(u_i/B)\right| \leq ((B-1)/B)\psi(p_{kB+1}). \tag{67}$$

Since $p_{(k+1)B} < \mu$, and since ψ is monotone increasing between 0 and μ, it follows that $\psi(p_{(k+1)B}) \geq \psi(p_{(k+1)B+1})$. So (65) gives

$$\begin{aligned}(1/B) \mid \psi(p_{kB+1}) + \cdots + \psi(p_{(k+1)B}) - B\,\psi(u_{k+1}/B) \mid \\ \leq ((B-1)/B)(\mathrm{MAX} - \psi(p_{kB+1})) \\ + ((B-1)/B)(\mathrm{MAX} - \psi(p_{(k+1)B+1})). \qquad (68)\end{aligned}$$

If we add together (66), (67), and (68), and use the triangle inequality, we obtain

$$(1/B)\left|\sum_{i=1}^{n} \psi(p_i) - B \sum_{i=1}^{n/B} \psi(u_i/B)\right| \leq (2(B-1)/B)\ \mathrm{MAX},$$

and as in the conclusion of case 1, this gives us (57).

It remains to prove (65). Assume that $1 \geq x_1 \geq \cdots \geq x_B \geq 0$. To prove (65), we must show

$$\begin{aligned}\mid ((\psi(x_1) + \cdots + \psi(x_B))/B \mid - \psi((x_1 + \cdots + x_B)/B) \mid \\ \leq ((B-1)/B)(\mathrm{MAX} - \psi(x_1)) \\ + ((B-1)/B)(\mathrm{MAX} - \psi(x_B)). \qquad (69)\end{aligned}$$

Assume that $\psi(x_B) \geq \psi(x_1)$. The proof is similar if $\psi(x_1) \geq \psi(x_B)$. Since $\psi(x_B) \geq \psi(x_1)$, clearly $\psi(x_i) \geq \psi(x_1)$, $i = 1, \cdots, B$

$$\begin{aligned}(\psi(x_1) + \cdots + \psi(x_B))/B - \psi((x_1 + \cdots + x_B)/B) \\ \leq ((B-1)\,\mathrm{MAX} + \psi(x_1))/B - \psi(x_1) \\ = ((B-1)/B)\,(\mathrm{MAX} - \psi(x_1)) \\ \leq ((B-1)/B)\,(\mathrm{MAX} - \psi(x_1)) \\ + ((B-1)/B)\,(\mathrm{MAX} - \psi(x_B)). \qquad (70)\end{aligned}$$

And,

$$\begin{aligned}\psi((x_1 + \cdots + x_B)/B) - (\psi(x_1) + \cdots + \psi(x_B))/B \\ \leq \mathrm{MAX} - ((B-1)\psi(x_1) + \psi(x_B))/B \\ \leq \mathrm{MAX} - ((B-1)\psi(x_1) + \psi(x_B))/B \\ + ((B-2)/B)\,(\mathrm{MAX} - \psi(x_B)) \\ = ((B-1)/B)\,(\mathrm{MAX} - \psi(x_1)) \\ + ((B-1)/B)\,(\mathrm{MAX} - \psi(x_B)). \qquad (71)\end{aligned}$$

Putting together (70) and (71), we obtain (69).

Remark The result of step 3 can be improved (reducing the right-hand side of (57) by less than a factor of 2), by taking advantage of more properties of ψ than that it is monotone increasing and then monotone decreasing.

11. *Proof of Step* 4

We will prove

$$0 \leq \sum_{i=1}^{n/B} u_i\,(1 - u_i/B)^T - \sum_{i=1}^{n B} (1 - u_i)^{T/B} < .271\,(B-1)/T + 33\,(B/T)^2 \qquad (72)$$

if $T \geq 2B$.

Statement (26) of the proof of step 2 says $(1 - u_i/B)^T \geq (1 - u_i)^{T/B}$. Hence $u_i\,(1 - u_i/B)^T \geq u_i\,(1 - u_i)^{T/B}$, and the first inequality of (72) follows.

If we adopt the notation of the proof of step 2, then we are concerned with finding $\sum_{i=1}^{n/B} u_i \epsilon_{T/B}(u_i)$. As before, if $\frac{1}{2} < u_i \leq 1$, then $i = 1$ and

$$u_i \epsilon_{T/B}(u_i) \leq (1 - 1/2B)^T < 8\,(B/T)^2, \quad \text{by (50) with } m = 2. \qquad (73)$$

If $0 \leq u_i \leq \frac{1}{2}$, then from (41) of step 2,

$$u_i \epsilon_{T/B}(u_i) \leq u_i^3\,(1 - u_i/B)^{T-B}\,((T(B-1)/2B^2) + .54). \qquad (74)$$

So from (73) and 74),

$$\sum_{i=1}^{n/B} u_i \epsilon_{T/B}(u_i) < 8\,(B/T)^2 + \sum_{i=1}^{n/B} u_i{}^3(1 - u_i/B)^{T-B}\,((T(B-1)/2B^2) + .54). \quad (75)$$

By exactly the same method as in step 2, where we now let ϕ be the function $x \to x^3(1 - x/B)^{T-B}$, we find that

$$\begin{aligned}\sum u_i{}^3(1 - u_i/B)^{T-B} &\leq (\max_{0\leq x\leq 1}(\phi(x)/x)) + (\max_{0\leq x\leq 1}\phi(x)) \\ &\leq 4B^2/e^2(T - B)^2 + 27B^2/e^3(T - B)^3,\end{aligned} \quad (76)$$

since (48) gives the maxima.

If we substitute into (75) using (76), we obtain

$$\begin{aligned}\sum_{i=1}^{n/B} u_i \epsilon_{T/B}(u_i) &< 8\left(\frac{B}{T}\right)^2 + \left(\frac{4B^2}{e^2(T-B)^2} + \frac{27B^3}{e^3(T-B)^3}\right)\left(\frac{T(B-1)}{2B^2} + .54\right) \\ &\leq 8\left(\frac{B}{T}\right)^2 + \frac{2(B-1)}{e^2T}\left(\frac{T}{T-B}\right)^2 + \frac{27}{2e^3}\left(\frac{B}{T-B}\right)^2\frac{T}{T-B} \\ &\quad + \frac{2.16}{e^2}\left(\frac{B}{T-B}\right)^2 + \frac{14.58}{e^3}\left(\frac{B}{T-B}\right)^3\end{aligned} \quad (77)$$

where we expanded out and made one substitution of B for $B - 1$.

If we now use (54) and (55) to substitute $2B/T$ for $B/(T - B)$ and $(1 + 2B/T)$ for $T/(T - B)$ in (77), we obtain

$$\begin{aligned}\sum_{i=1}^{n/B} u_i \epsilon_{T/B}(u_i) &< 8\left(\frac{B}{T}\right)^2 + \frac{2(B-1)}{e^2T}\left(1 + \frac{2B}{T}\right)^2 + \frac{27}{2e^3}\left(\frac{2B}{T}\right)^2\left(1 + \frac{2B}{T}\right) \\ &\quad + 2.16\left(\frac{2B}{T}\right)^2 + \frac{14.58}{e^3}\left(\frac{2B}{T}\right)^3 \\ &\leq \frac{2(B-1)}{e^2T} + \left(\frac{B}{T}\right)^2\left[8 + \frac{8}{e^2} + \frac{54}{e^3} + 8.64\right] \\ &\quad + \left(\frac{B}{T}\right)^3\left[\frac{8}{e^2} + \frac{108}{e^3} + \frac{116.64}{e^3}\right],\end{aligned} \quad (78)$$

where we expanded and occasionally substituted B for $B - 1$.

If we replace $(B/T)^3$ on the right-hand side of (71) by $(B/T)^2$ and numerically evaluate, we obtain

$$\sum_{i=1}^{n/B} u_i \epsilon_{T/B}(u_i) < .271\,(B-1)/T + 32.68\,(B/T)^2,$$

which implies the second inequality of (72).

12. *Proof of Step* 5

We will prove that

$$H(T)/H(t) \leq S(T)/S(t) \quad (79)$$

if $T \geq t \geq 1$.

We will show that $H(T)/S(T) \leq H(t)/S(t)$, that is,

$$\begin{aligned}\left(1 - \sum_{i=1}^{n} p_i(1-p_i)^T\right)\Big/\sum_{i=1}^{n}(1 - (1-p_i)^T) \\ \leq \left(1 - \sum_{i=1}^{n} p_i(1-p_i)^t\right)\Big/\sum_{i=1}^{n}(1 - (1-p_i)^t).\end{aligned} \quad (80)$$

Assume for convenience that each p_i is nonzero.

Let $\alpha_i = 1 - (1 - p_i)^t$ $\beta_i = 1 - (1 - p_i)^T$, $1 \leq i \leq n$. Then (80) is equivalent to

$$\sum_{i=1}^{n} p_i\beta_i \Big/ \sum_{i=1}^{n} \beta_i \leq \sum_{i=1}^{n} p_i\alpha_i \Big/ \sum_{i=1}^{n} \alpha_i. \tag{81}$$

Let $\alpha_i' = \alpha_i / \sum_{j=1}^{n} \alpha_j$, $\beta_i' = \beta_i / \sum_{j=1}^{n} \beta_j$, $1 \leq i \leq n$. Then $\sum \alpha_i' = \sum \beta_i' = 1$. Statement (81) is equivalent to

$$\sum p_i\beta_i' \leq \sum p_i\alpha_i'. \tag{82}$$

Now the function f, given by $\mathbf{x} \to \sum_{i=1}^{n} p_i x_i$, with domain n-tuples $(x_1, \cdots, x_n)$ with $x_1 \geq x_2 \geq \cdots \geq x_n$, is a Schur function, since $(\partial/\partial x_i)f = p_i$, $1 \leq i \leq n$, and hence $(\partial/\partial x_i)f > (\partial/\partial x_j)f$ *iff* $p_i > p_j$ *iff* $i > j$ *iff* $x_i > x_j$. So, to prove (82), we need only show

$$\boldsymbol{\alpha}' \succ \boldsymbol{\beta}'. \tag{83}$$

To prove (83), we need the following lemma.

LEMMA. *If* $0 \leq r_1 \leq r_2 < 1$, *then*

$$(1 - r_1^t)/(1 - r_2^t) \geq (1 - r_1^T)/(1 - r_2^T). \tag{84}$$

PROOF. It is easy to see that (84) holds if $r_1 = 0$; hence, assume $r_1 > 0$. Let f_1 be the function $x \to (1 - r_1^x)/(1 - r_2^x)$, with domain $[1, \infty)$. We need only show that f_1 is monotone decreasing, that is, that $f_1' \leq 0$. A simple calculation shows that $f' \leq 0$ iff

$$r_1^x \log r_1/(1 - r_1^x) \geq r_2^x \log r_2/(1 - r_2^x) \quad \text{for each } x \geq 1. \tag{85}$$

If we multiply both sides of (85) by x, and let $g_1 = r_1^x$, $g_2 = r_2^x$, we obtain

$$g_1 \log g_1/(1 - g_1) \geq g_2 \log g_2/(1 - g_2). \tag{86}$$

So we need only show that (86) holds whenever $0 \leq g_1 \leq g_2 < 1$. Let $s_1 = 1 - g_1$, $s_2 = 1 - g_2$. Then (86) becomes

$$(1 - s_1)\log(1 - s_1)/s_1 \geq (1 - s_2)\log(1 - s_2)/s_2. \tag{87}$$

We need only show that (87) holds whenever $0 < s_2 \leq s_1 < 1$. Let f_2 be the function $x \to (1 - x)\log(1 - x)/x$, with domain $(0, 1)$. We must show that f_2 is monotone increasing. If we replace $\log(1 - x)$ by its Taylor expansion, then

$$\begin{aligned} f(x) &= (1 - x)\left(-x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots\right)\Big/ x \\ &= -(1 - x)\left(1 + \frac{x}{2} + \frac{x^2}{3} + \frac{x^3}{4} + \cdots\right) \\ &= -1 + \left(1 - \frac{1}{2}\right)x + \left(\frac{1}{2} - \frac{1}{3}\right)x^2 + \left(\frac{1}{3} - \frac{1}{4}\right)x^3 + \cdots, \end{aligned}$$

which is clearly monotone increasing. The lemma is now proved.

To prove (83), we must show

$$\sum_{i=1}^{k} \alpha_i' \geq \sum_{i=1}^{k} \beta_i', \qquad 1 \leq k < n. \tag{88}$$

(We have already noted that $\sum_{i=1}^{n} \alpha_i' = \sum_{i=1}^{n} \beta_i' = 1$.)

We will first prove $\alpha_1' \geq \beta_1'$. In the lemma, if we let $r_1 = 1 - p_i$, $r_2 = 1 - p_{i+1}$, then we get

$$\alpha_i/\alpha_{i+1} \geq \beta_i/\beta_{i+1}, \qquad 1 \leq i < n. \tag{89}$$

We will now show that (89) implies

$$\alpha_i' \geq \beta_i'. \tag{90}$$

For, find $\sigma_1, \cdots, \sigma_{n-1}$ and $\tau_1, \cdots, \tau_{n-1}$ such that $\alpha_{i+1} = \sigma_i \alpha_i$, $\beta_{i+1} = \tau_i \beta_i$, $1 \leq i < n$. Then (89) implies

$$\tau_i \geq \sigma_i, \qquad 1 \leq i < n. \tag{91}$$

Now

$$\begin{aligned} \alpha_1' &= \alpha_1/(\alpha_1 + \cdots + \alpha_n) \\ &= \alpha_1/\alpha_1(1 + \sigma_1 + \sigma_1\sigma_2 + \sigma_1\sigma_2\sigma_3 + \cdots) \\ &= 1/(1 + \sigma_1 + \sigma_1\sigma_2 + \sigma_1\sigma_2\sigma_3 + \cdots). \end{aligned} \tag{92}$$

Similarly

$$\beta_1' = 1/(1 + \tau_1 + \tau_1\tau_2 + \tau_1\tau_2\tau_3 + \cdots). \tag{93}$$

Then (91), (92), and (93) imply (90).

We will close by showing that $\alpha_1' + \alpha_2' \geq \beta_1' + \beta_2'$. The other inequalities of (88) are proved very similarly.

Let $\hat{\alpha} = (\alpha_1 + \alpha_2, \alpha_3, \alpha_4, \cdots, \alpha_n)$ and $\hat{\beta} = (\beta_1 + \beta_2, \beta_3, \beta_4, \cdots, \beta_n)$. We will first prove that

$$\hat{\alpha}_i/\hat{\alpha}_{i+1} \geq \hat{\beta}_i/\hat{\beta}_{i+1} \tag{94}$$

for each i. This follows for $i \geq 2$ by (89). If $i = 1$, then (94) says

$$(\alpha_1 + \alpha_2)/\alpha_3 \geq (\beta_1 + \beta_2)/\beta_3. \tag{95}$$

But (95) holds, since $\alpha_1/\alpha_3 \geq \beta_1/\beta_3$ and $\alpha_2/\alpha_3 \geq \beta_2/\beta_3$, each by an application of the lemma, just as in the proof of (89).

Let $\hat{\alpha}'$ be the normalization of $\hat{\alpha}$, that is, $\hat{\alpha}'_i = \hat{\alpha}_i / \sum_{j=1}^{n} \hat{\alpha}_j$. Similarly, define $\hat{\beta}'$. (94) implies

$$\hat{\alpha}'_1 \geq \hat{\beta}'_1, \tag{96}$$

just as (89) implies (90).

But (96) is equivalent to $\alpha_1' + \alpha_2' \geq \beta_1' + \beta_2'$, which was to be shown.

13. *Summary*

We have shown that the working-set miss ratio is insensitive to page size, in the independent (page) reference model. We argue that this implies that the insensitivity also holds in more realistic models.

ACKNOWLEDGMENT. The authors wish to thank A.C. McKellar for very helpful discussions about how our results relate to actual computing systems.

REFERENCES

(Note. References [8, 14] are not cited in the text.)

1. AHO, A V., DENNING, P J., AND ULLMAN, J.D. Principles of optimal page replacement *J. ACM 18*, 1 (Jan. 1971), 80–93
2. ANACKER, W , AND WANG, C.P. Performance evaluation of computing systems with memory hierarchies *IEEE Trans. Electronic Computers EC-16* (1967), 765–773.
3. BENNETT, B.T. Private communication
4. CHU, W.W., AND OPDERBECK, H Performance of replacement algorithms with different page sizes. *Computer 7*, 11 (Nov 1974), 14–21.
5. DENNING, P.J., AND SCHWARTZ, S.C Properties of the working-set model. *Comm. ACM 15*, 3 (March 1972), 191–198.

3. DENNING, P J., SAVAGE, J E., AND SPIRN, J.R. Models for locality in program behavior Princeton U., Princeton, N.J., April 1972
   EASTON, M.C. Model for interactive data base reference strings *IBM J. Res. Develop. 19*, 6 (1975), 550–556.
   EASTON, M.C , AND BENNETT, B T. On transient-free working-set statistics Res Rep. RC 5140, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y , Nov. 1974.
   EASTON, M.C., AND FAGIN, R. Cold-start vs. warm-start miss ratios and multiprogramming performance. (To appear).

10 FAGIN, R. A counterintuitive example of computer paging. Res. Rep. RC 5031, IBM Thomas J. Watson Research Center, Yorktown Heights, N Y , Aug. 1974; *Comm ACM* (to appear)
11. FAGIN, R. Asymptotic miss ratios over independent references Res. Rep. RC 5415, IBM Thomas J Watson Research Center, Yorktown Heights, N Y., May 1975.
12 GELENBE, E , TIBERIO, P , AND BOEKHOERST, J C A Page size in demand paging systems *Acta Informatica 3,* 1 (1973), 1–23
13 HARDY, G.H , LITTLEWOOD, J.E , AND POLYA, G *Inequalities.* Cambridge U Press, London, 1964
14. KING, W.F III Analysis of paging algorithms IFIP Conf Proc , Ljubljana, Yugoslavia, Aug 1971.
15. KNOPP, K *Infinite Sequences and Series.* Dover, New York, 1956.
16 KUCK, D J , AND LAWRIE, D H The use and performance of memory hierarchies. A survey. In *Computer and Information Sciences II,* J.T. Tou, Ed , Academic Press, New York, 1969
17 KAPLAN, K R , AND WINDER, R.D. Cache-based computer systems. *Computer 6,* 3 (March 1973), 30–36
18. LEWIS, P.A W , AND SHEDLER, G.S Empirically derived micromodels for sequences of page exceptions *IBM J. Res. Develop. 17,* 2 (March 1973), 86–100
19 MATTSON, R., GECSEI, J , SLUTZ, D , AND TRAIGER, I. Evaluation techniques for storage hierarchies. *IBM Syst. J 9,* 2 (1970), 78–117
20 MARSHALL, A.W , OLKIN, I , AND PROSCHAN, F. Monotonicity of ratios of means and other applications of majorization. In *Inequalities,* D. Shisha, Ed., Academic Press, New York, 1967
21 SCHUR, I Über ein Klasse von Mittlebildungen mit Anwendungen auf die Determinatentheorie *Sitzber. Berl. Math. Ges 22* (1923), 9–20.
22. WIMBROW, J H A large-scale interactive administrative system *IBM Syst. J 10,* 4 (1971), 260–282
23 YUE, P C., AND WONG, C K. On the optimality of the probability ranking scheme in storage applications. *J ACM 20,* 4 (Oct 1973), 624–633

RECEIVED OCTOBER 1974, REVISED JULY 1975