JOURNAL OF COMPUTER AND SYSTEM SCIENCES 14, 222–250 (1977)

Asymptotic Miss Ratios over Independent References

RONALD FAGIN*

IBM Research Laboratory, San Jose, California 95193

Received June 18, 1975; revised July 14, 1976

It is proven that under the assumption of independent references, the (apparently analytically and computationally intractable) expected LRU (least recently used) miss ratio with main memory size CAP can be approximated arbitrarily closely by the (analytically and computationally tractable) expected working-set miss ratio with expected working-set size CAP, as the size of the database goes to infinity. Their common asymptotic value is given by a tractable formula involving integrals. An immediate corollary of the representation is the asymptotic independence of miss ratio from page size in the independent reference model and in some generalizations of this model. This result also has implications about the effect on miss ratio of variable or fixed partitioning of main memory, in case of multiprogramming. Furthermore, in certain database environments, we can answer the question as to how the size of main memory must vary in order to maintain the same miss ratio, when the size of the database increases. The methods of this paper are extended to give an asymptotic formula for the miss ratio under VMIN, the optimal variable-space page replacement algorithm under demand paging.

1. INTRODUCTION

A major difficulty in the mathematical analysis of computer system performance is the lack of analytically tractable formulas for page fault behavior in even some very simple models of page reference patterns. In this paper, we will give a solution to this problem in the case of the well-known LRU (least recently used) miss ratio [1, 15] in the simple, widely studied "independent reference model" [1, 4, 9, 12] in a paged, two-level storage hierarchy. The solution consists of replacing the LRU miss ratio formula by a tractable formula which we prove is close to the LRU miss ratio in this model for large database sizes.

There is another way to view the results in this paper. An important question about a multiprogrammed computing system is how to divide the main (first-level) memory among competing programs. One issue is whether there should be a fixed or a variable partition of main memory. When each program is operated under LRU memory management, then we have one example of a fixed partition of memory; when each program is operated under a working-set memory management policy, then we have a variable partition. So an analysis of the LRU vs the working-set miss ratio gives a comparison

* This research was conducted while the author was at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, 10598.

ISSN 0022-0000

between one natural fixed-partition policy and one natural variable-partition policy. Several papers [3, 16] have dealt with this very comparison.

In this paper, we show that in the independent reference model, there is little difference in the miss ratio for the two cases (in fact, under suitable assumptions, as the number of pages gets large, the miss ratios get arbitrarily close). Further, we show that this indifference also holds in a page reference model (due to Easton [5]), which is a generalization of the independent reference model, but in which there is locality of reference.

Throughout this paper, we analyze miss-ratio behavior as the number of pages goes to infinity ("asymptotic behavior"). Actually, for each of the three applications of our results which are discussed in this paper, we are interested in the miss-ratio behavior when the number of pages is fixed and "sufficiently large." In the first (and most important) application, we use the fact that under suitable assumptions, the limit (as the number of pages goes to infinity) of the expected LRU miss ratio and the limit of the corresponding expected working-set miss ratio converge to the same value; because of this fact, we know that if the number of pages is large (but fixed), then the expected LRU and expected working-set miss ratios are approximately equal. Furthermore, our asymptotic formula gives a value which is close to the correct expected miss ratio when the number of pages is large but fixed. In the second application (in Section 5), we show that in the independent reference model (and some generalizations), if we hold fixed both the total number of bytes in the database and the number of bytes in main memory, but double the number of pages by halving the page size, then the expected miss ratio remains approximately constant. In the third application (in Section 6), we show that under not unreasonable assumptions, if we know that our database will double in size in the next five years, then to maintain the same miss ratio we must approximately double the size of main memory.

In the *independent reference model*, we assume that at each discrete time t ($t = 1, 2, 3,...$), one page is referenced, where page i is referenced with probability p_i, independent of past history. Of course, $\sum p_i = 1$. Assume that the *capacity*, or size of main memory, is CAP pages. Under the LRU memory management policy, if there is a page fault, that is, if a page is referenced that is not in main memory, then that page is moved into main memory, and the page that has been least recently referenced is removed. It is easy to see that main memory always contains the CAP pages which have been most recently referenced. The *expected LRU miss ratio* is the limiting probability of a page fault, that is, the limit (as $T \to \infty$) of the probability that the Tth page reference is a page fault. It is well known that the limit exists, and is independent of the initial configuration of main memory.

King [12] derived the following formula for the expected LRU miss ratio (in the independent reference model):

$$\sum \frac{p_{i_1} p_{i_2} \cdots p_{i_{\mathrm{CAP}}}(1 - p_{i_1} - \cdots - P_{i_{\mathrm{CAP}}})}{(1 - p_{i_1})(1 - p_{i_1} - p_{i_2}) \cdots (1 - p_{i_1} - p_{i_2} - \cdots - p_{i_{\mathrm{CAP}-1}})}, \tag{1.1}$$

where the sum is taken over all CAP-tuples ($i_1, ..., i_{\mathrm{CAP}}$) such that $i_j \neq i_k$ if $j \neq k$. This formula is sufficiently complex that it seems difficult to derive or prove interesting analytic results by using it. In addition, the formula contains so many terms (roughly n^{CAP}, where n is the number of pages and CAP is the capacity), that it is impossible to numerically evaluate it (in the straightforward manner) for moderate size n and CAP except in the most trivial cases (such as $p_1 = \cdots = p_n = 1/n$, where the formula collapses). We remark that Lenfant [14] has obtained another formula for the expected LRU miss ratio in the independent reference model. Lenfant's formula contains $(n\text{-CAP})2^n$ terms, and it also seems analytically and computationally intractable.

The *expected working-set miss ratio* (*with window size* T) [4] in the independent reference model is defined to be the probability that the page referenced at time $t > T$ was not one of the pages referenced over the course of the previous T (not necessarily distinct) references. It is easy to see that this value is independent of t, for $t > T$, and that the value is

$$M(T) = \sum_{i=1}^{n} p_i(1 - p_i)^T, \tag{1.2}$$

if there are n pages with reference probabilities $p_1, \ldots, p_n$. For, $p_i(1 - p_i)^T$ is the probability that page i is the next page referenced, and that page i was not referenced during the previous T references. Another quantity of interest, the expected number of distinct pages to be referenced over the course of T references, or the *expected working-set size*, is

$$S(T) = \sum_{i=1}^{n} (1 - (1 - p_i)^T), \tag{1.3}$$

since $1 - (1 - p_i)^T$ is the probability that page i was referenced. Thus, if $S(T) = \mathrm{CAP}$, then the expected working-set miss ratio with expected working-set size CAP is $M(T) = M(S^{-1}(\mathrm{CAP}))$. Note that $M(S^{-1}(\mathrm{CAP}))$ is well defined, via (1.2) and (1.3), for all real CAP between 0 and n, even if the intermediate parameter $T = S^{-1}(\mathrm{CAP})$ is not an integer: This procedure gives us a convenient interpolation.

Now $M(S^{-1}(\mathrm{CAP}))$ is easy to evaluate numerically: First, find T by binary search such that $S(T) = \mathrm{CAP}$, and then find $M(T)$ for this T. In addition, this formula is analytically tractable. For example, from this formula upper bounds are obtained in [7] for the effect of page size on the working-set miss ratio. There appears to be no way to obtain similar results about the LRU miss ratio by using King's formula.

In this paper, we show that the expected working-set miss ratio with expected working-set size CAP is close to the LRU miss ratio with size of main memory CAP in a certain precise asymptotic sense. In addition, we exhibit a formula, involving integrals, for their common asymptotic value. This formula depends on two factors: The shape of the cumulative probability distribution function (such as Zipf's law with a given skewness), and the fraction of pages which can fit in main memory. This formula, like the formula for the expected working-set miss ratio with expected working-set size CAP, is analytically and computationally tractable. We also give a formula (Section 4) for the asymptotic value of the expected VMIN miss ratio, where VMIN is the optimal variable-space page replacement algorithm under demand paging [17, 18].

We close this section with a brief digression on the independent reference model. The assumption that page references are independent can be justified in some cases, such as for archival store references [20]. In other cases, such as for program traces, there is a great deal of serial correlation between page references, and hence the independence assumption fails badly. However, even here, where the independent reference model does not apply, results about the independence reference model can yield interesting conclusions. For example, in [7] an argument is presented that if the LRU miss ratio is approximately independent of page size in the independent reference model, then the LRU miss ratio is approximately independent of page size in more realistic models also, in which page references have an "independent" component and a "sequential" component. Finally, as Gelenbe observes [9], results about the independent reference model have a more universal flavor than results about more realistic but more restricted reference processes.

2. Summary of Results

Since we wish to discuss asymptotic behavior as the number of pages gets large, we need a canonical method for determining a probability distribution $\{p_1^{(n)},..., p_n^{(n)}\}$, for each n. We proceed as follows.

Let F be a smooth,[1] monotone increasing function with domain the closed interval $[0, 1]$, such that $F(0) = 0$ and $F(1) = 1$. We will call F a *cumulative probability distribution function*. For each positive integer n, we can define a probability distribution $\{p_1^{(n)},..., p_n^{(n)}\}$ by setting

$$p_i^{(n)} = F(i/n) - F((i-1)/n), \qquad 1 \leqslant i \leqslant n. \tag{2.1}$$

When n is fixed, we may write p_i for $p_i^{(n)}$. Of course, $p_i \geqslant 0$ for each i, and $\sum p_i = 1$. We will call $\{p_1,..., p_n\}$ the *probability distribution determined by F (and n)*. In Fig. 1, we sketch the situation for $n = 4$.

Caution. The reader should *not* assume that going from, say, $n = 4$ to $n = 8$ means that we have doubled the number of pages by halving the page size, with the total number of bytes in the database remaining fixed. This will be the correct scenario for only one application of our results (the independence of miss ratio from page size, Section 5); it will not be the correct scenario in general. A better viewpoint is that in going from $n = 4$ to $n = 8$, we have doubled the number of pages while leaving the page size fixed, which means that the total number of bytes in the database has doubled. Because the number of pages has changed (from 4 to 8), there must be a new probability distribution, and we have given a formula, or mechanism, for determining this probability distribution, through using a cumulative probability distribution function F which is assumed to always remain fixed.

Let β_0 be a real number, $0 \leqslant \beta_0 \leqslant 1$. Intuitively, β_0 will be CAP/n, where CAP is the

[1] By "smooth," we mean continuously differentiable. To include the important Zipf's law case (Section 3), we will allow the possibility that $F'(0) = \infty$.

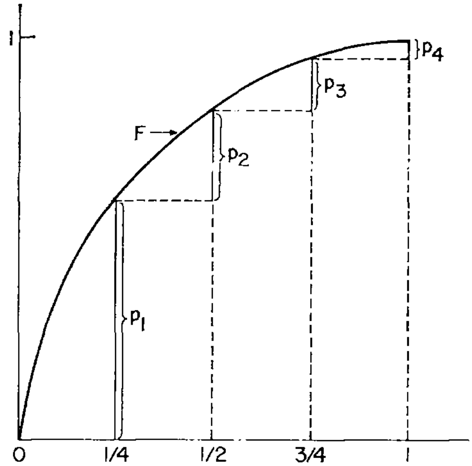

FIGURE 1

capacity, or size of first-level memory (in pages) in the LRU case and the expected working-set size (in pages) in the working-set case, and where n is the total number of pages.

Denote by LRU(n, β_0) the expected LRU miss ratio in the independent reference model with probability distribution $\{p_1, ..., p_n\}$ determined by F, and with capacity $\lfloor \beta_0 n \rfloor$. (By $\lfloor x \rfloor$, we mean the greatest integer not exceeding x; similarly, $\lceil x \rceil$ is the least integer not less than x.) We will show that LRU(n, β_0) converges to a value MISS(β_0) as $n \to \infty$. This result says that if the cumulative probability distribution function F and the fraction β_0 of pages which can fit in main memory are each held fixed, then the expected LRU miss ratio in the independent reference model converges, as the number n of pages goes to infinity.

Let WORK(n, β_0) be the expected working-set miss ratio with expected working-set size $\lfloor \beta_0 n \rfloor$. We will show that WORK$(n, \beta_0)$ converges as $n \to \infty$, to the same value MISS(β_0). This result says that if the cumulative probability distribution function F and the ratio β_0 of the expected working-set size divided by the total number of pages are each held fixed, than the expected working-set miss ratio in the independent reference model converges, as the number n of pages goes to infinity, and that the limiting value is the same as in the LRU case. In particular, WORK(n, β_0) and LRU(n, β_0), the working-set and LRU miss ratios, are close for large n.

We will now explicitly define the limiting value MISS(β_0). Let τ_0 and μ_0 be new parameters, where $0 \leqslant \tau_0 \leqslant \infty$, and $0 \leqslant \mu_0 \leqslant 1$. Intuitively, τ_0 will be T/n, where T is the window size and n is the number of pages. The parameter μ_0 will be the limiting miss ratio MISS(β_0).

We will demonstrate a natural one-one correspondence between β_0's and τ_0's, given by

$$\beta_0 = 1 - \int_0^1 e^{-\tau_0 F'(x)}\, dx. \tag{2.2}$$

The correspondence is one–one, since if $\tau_0 \to 0$, then $\beta_0 = 0$, and if $\tau_0 \to \infty$, then $\beta_0 = 1$; further, the right-hand side of (2.2) is a strictly monotone-increasing function of τ_0.

There will also be a natural one–one correspondence between τ_0's and μ_0's, given by

$$\mu_0 = \int_0^1 F'(x)\, e^{-\tau_0 F'(x)}\, dx. \tag{2.3}$$

To find MISS(β_0), first find τ_0 such that (2.2) holds. Then let MISS(β_0) $= \mu_0$, where μ_0 relates to τ_0 via (2.3). In other words, define functions β^* and μ^* as

$$\begin{aligned} \beta^*(\tau_0) &= 1 - \int_0^1 e^{-\tau_0 F'(x)}\, dx, \\ \mu^*(\tau_0) &= \int_0^1 F'(x)\, e^{-\tau_0 F'(x)}\, dx. \end{aligned} \tag{2.4}$$

Then

$$\mathrm{MISS}(\beta_0) = \mu^*(\beta^{*-1}(\beta_0)). \tag{2.5}$$

Note from (2.4) that

$$d\beta^*/d\tau_0 = \mu^*, \tag{2.6}$$

since we can differentiate under the integral sign by continuity of the integrands in (2.4) [10, p. 106]. Equation (2.6) is the continuous analog of Denning and Schwartz's difference equation

$$S(T+1) - S(T) = M(T). \tag{2.7}$$

Equation (2.7) can be verified directly (for the independent reference model) from (1.2) and (1.3).

We close this section with a more detailed explanation of the sense in which the "normalized capacity" β_0, the "normalized window-size" τ_0, and the limiting miss ratio μ_0 correspond. Define

$$\begin{aligned} \mu(n, \tau_0) &= \sum_{i=1}^{n} p_i^{(n)}(1 - p_i^{(n)})^{\tau_0 n}, \\ \beta(n, \tau_0) &= 1 - (1/n) \sum_{i=1}^{n} (1 - p_i^{(n)})^{\tau_0 n}, \end{aligned} \tag{2.8}$$

where $\{p_1^{(n)}, \ldots, p_n^{(n)}\}$ is the probability distribution determined by F and n. Comparing Eqs. (2.8) with (1.2) and (1.3), we see that if $T = \tau_0 n$ is an integer, then $\mu(n, \tau_0)$ is the expected working-set miss ratio, and $\beta(n, \tau_0)$ the ratio of the expected working-set size divided by the number of pages, when T is the window size (if T is not an integer, then these are interpolated values). We will show that

$$\begin{aligned} \mu(n, \tau_0) &\to \mu^*(\tau_0) = \mu_0, && \text{as } n \to \infty, \\ \beta(n, \tau_0) &\to \beta^*(\tau_0) = \beta_0, && \text{as } n \to \infty. \end{aligned}$$

3. Special Cases

Zipf's Law. G. K. Zipf found that many naturally occurring probability distributions follow "Zipf's law" [21; 13, p. 397], in which the probability z_i that page i is referenced is

$$z_i = k/i^{\theta}, \qquad 1 \leqslant i \leqslant n,$$

where θ is a positive constant (the "skewness"), and k is a normalizing constant chosen so that $\sum z_i = 1$. For example, for the probability distribution of words in natural language texts, Zipf found that $\theta \approx 1$, and for the distribution of personal income, $\theta \approx 0.5$. The well-known "80/20 law" [11], which states that 80% of the references to a file occur to only 20% of the file (and that 80% of these references occur to only 20% of the top 20%, and so on), can be approximated by a Zipf's law distribution with $\theta \approx 0.86$ [13, p. 398].

It would be nice if for each skewness θ, there was an easily evaluable function ZIPF $=$ ZIPF$_\theta$ such that ZIPF(β_0) were the limit (as the number of pages goes to infinity) of the expected LRU miss ratio for a Zipf's law distribution when the fraction β_0 of pages can fit in main memory.

If there was a cumulative probability distribution function $F = F_\theta$ such that the Zipf's law distribution with n pages (and skewness θ) were exactly the probability distribution determined by F and n as before, then there would be such a function ZIPF: Namely, the function MISS $=$ MISS$_F$ defined in Eq. (2.5). We will show in Appendix 1 that the function $F: x \to x^{1-\theta}$ "almost" determines the Zipf's law distribution with skewness θ (for $0 \leqslant \theta < 1$), and in Appendix 2 that this is good enough, that is, that ZIPF$_\theta$ $=$ MISS$_F$. By "almost" in the previous sentence, we mean that if $\{z_1^{(n)},..., z_n^{(n)}\}$ is the Zipf's law distribution with n pages (and with skewness θ), and if β_0 lies between 0 and 1, then

$$\sum_{i=1}^{\lfloor \beta_0 n \rfloor} z_i^{(n)} \to F(\beta_0), \qquad \text{as} \quad n \to \infty. \tag{3.1}$$

That is, if $0 \leqslant \theta < 1$, then (3.1) holds, where $F(\beta_0) = \beta_0^{1-\theta}$.

We can now obtain the asymptotic value (as the number of pages gets large) of the expected LRU miss ratio in the independent reference model with a Zipf's law distribution with skewness θ, when $0 \leqslant \theta < 1$. Let β_0 be the fraction of pages which can fit in main memory. Then the limiting expected LRU miss ratio is MISS$(\beta_0) = \mu^*(\beta^{*-1}(\beta_0))$, where

$$\beta^*(\tau_0) = 1 - \int_0^1 e^{-\tau_0(1-\theta)x^{-\theta}}\, dx, \qquad 0 \leqslant \tau_0 < \infty, \tag{3.2}$$

$$\mu^*(\tau_0) = \int_0^1 (1-\theta)\, x^{-\theta} e^{-\tau_0(1-\theta)x^{-\theta}}\, dx, \qquad 0 \leqslant \tau_0 < \infty. \tag{3.3}$$

To calculate MISS(β_0) numerically, find τ_0 by binary search such that $\beta^*(\tau_0) = \beta_0$, and then find $\mu_0 = \mu^*(\tau_0)$, the limiting miss ratio.

Let us give a numerical example.[2] We will round off all results to four decimal places. Let the skewness θ be 0.5, and let β_0 be 0.6. From Eq. (3.2), we find that $\beta_0 = 0.6$ corresponds to $\tau_0 = 1.1403$ (that is, $\beta^*(1.1403) = 0.6$). From (3.3), we find that $\mu^*(1.1403) = 0.2902$. So the limiting value MISS(0.6) is 0.2902. Thus, with a Zipf's law distribution with skewness $\theta = 0.5$, if 60% of the pages can fit in main memory, then the expected LRU miss ratio converges to $\mu_0 \approx 0.2902$ as the number of pages gets large. In addition, the expected working-set miss ratio, where the expected working-set size is 60% of the number of pages, converges to μ_0. As an empirical confirmation, we can calculate the expected working-set miss ratio for various values of n, when the expected working-set size CAP is 60% of n (See Table I.)

TABLE I[a]

Zipf's Law, Skewness $\theta = 0.5$ and $\beta_0 = 0.6$

n	CAP	Working-set miss ratio	LRU miss ratio
10	6	0.3492	0.3518
100	60	0.3109	?
1000	600	0.2968	?
10000	6000	0.2923	?
Limiting value		0.2902	0.2902

[a] Values rounded to four decimal places.

In the case of LRU, if $n = 10$ and CAP $= 6$, then King's formula has over 30,000 terms and involves over 300,000 multiplications and divisions. When $n = 100$ and CAP $= 60$, then King's formula has $(100!)/(40!) \approx 10^{110}$ terms, so of course direct computation is out of the question.

What about the case of Zipf's law when $\theta \geqslant 1$? We remark that in this case, there is a limiting miss ratio function ZIPF_θ as described earlier, and that it is degenerate: $\mathrm{ZIPF}_\theta(\beta_0)$ is 1 if $\beta_0 = 0$, and 0 otherwise.

Arithmetic Probability Distribution

Assume that a and b are constants (a nonnegative and b positive), and that

$$p_i^{(n)} = k(a + ib), \qquad 1 \leqslant i \leqslant n,$$

where k is a normalizing constant chosen so that $\sum_{i=1}^{n} p_i^{(n)} = 1$. Then we say that $\{p_1^{(n)},..., p_n^{(n)}\}$ is an arithmetic probability distribution.

It is straightforward to check that the quantity $p_1^{(n)} + \cdots + p_{\lfloor\beta_0 n\rfloor}^{(n)}$ converges to $\beta_0{}^2$ as $n \to \infty$, independent of a and b. Hence, all arithmetic probability distributions

[2] All calculations were carried out on the IBM 370/168 at the IBM Thomas J. Watson Research Center.

correspond to the cumulative probability distribution function $F: x \to x^2$, in the same approximate sense that Zipf's law with skewness θ, $0 \leqslant \theta < 1$, corresponds to $F: x \to x^{1-\theta}$. So, the results of Appendix 2 imply that the limit (as the number of pages goes to infinity) of the expected (LRU or working-set) miss ratio as a function of β_0 is $\mathrm{MISS}(\beta_0) = \mu^*(\beta^{*-1}(\beta_0))$, where

$$\begin{aligned} \beta^*(\tau_0) &= 1 - \int_0^1 e^{-2\tau_0 x}\,dx = \frac{2\tau_0 + e^{-2\tau_0} - 1}{2\tau_0}, \\ \mu^*(\tau_0) &= \int_0^1 2xe^{-2\tau_0 x} = \frac{1 - e^{-2\tau_0}(1 + 2\tau_0)}{2\tau_0^2}. \end{aligned} \tag{3.4}$$

In the next section, we will make some observations concerning the expected A_0 miss ratio and the expected VMIN miss ratio in the case of an arithmetic probability distribution.

4. Asymptotic Values of the Expected A_0 and VMIN Miss Ratios

A_0 is the optimal page replacement algorithm with no knowledge of the future in the independent reference model [1]. In the A_0 page replacement algorithm, if the capacity is CAP, then main memory always contains the (CAP $-$ 1) pages with the largest reference probabilities, along with the most recently referenced of the remaining (n $-$ CAP $+$ 1) pages. It is not hard to check that if F is concave (that is, if the derivative F' is monotone decreasing; this corresponds to $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$) as it is in the Zipf's law case, then the limit (as $n \to \infty$) of the expected A_0 miss ratio, when the fraction β_0 of pages can fit in main memory, is $1 - F(\beta_0)$; if F is convex, as it is in the arithmetic case, then the limit (as $n \to \infty$) of the expected A_0 miss ratio is $F(1 - \beta_0)$ (in the arithmetic case, F is convex, since for convenience we defined arithmetic probability distributions $\{p_1, \ldots, p_n\}$ in such a way that $p_1 \leqslant p_2 \leqslant \cdots \leqslant p_n$.)

VMIN [17, 18] is the optimal variable-space page replacement algorithm under demand paging. (By a *demand-paging* algorithm, we mean one in which a page may be brought into main memory only in the event of a page fault; however, a page may be removed from main memory at any time.) The VMIN algorithm, like the working-set algorithm, has associated with it a fixed window-size T. When a page is referenced, it is brought into main memory if it is not yet present. If the page is not rereferenced within the next T time units, then it is immediately removed from main memory just after it is referenced; otherwise, it is retained in main memory, at least until when it is first rereferenced. For a given value of T and a given page reference string, let M be the VMIN miss ratio, let VCAP be the average number of pages in main memory under the VMIN page replacement algorithm, and let WCAP be the average number of pages in main memory under the working-set page replacement algorithm (with the same window-size T.) Slutz shows [18] that

$$\mathrm{VCAP} = \mathrm{WCAP} - (T - 1)M. \tag{4.1}$$

It will be convenient to "normalize" Eq. (4.1) by dividing both sides by n (the number of pages), to obtain

$$\mathrm{VCAP}/n = (\mathrm{WCAP}/n) - ((T - 1)/n)M. \tag{4.2}$$

Let us now consider the independent reference model. Assume that we hold fixed a cumulative probability distribution function F. For each τ_0 $(0 \leqslant \tau_0 < \infty)$, define

$$\beta^{**}(\tau_0) = \beta^*(\tau_0) - \tau_0\mu^*(\tau_0), \tag{4.3}$$

where β^* and τ^* are defined in (2.4). From (4.2) and from the fact that the VMIN and working-set miss ratios with window-size T are equal (so M in (4.2) can be considered the working-set miss ratio), along with our discussion of β^* and μ^* at the end of Section 2, it follows that if a "normalized window-size" τ_0 is held fixed as in Section 2, then the expected value of VCAP/n converges (as $n \to \infty$) to $\beta^{**}(\tau_0)$, and the expected value of the VMIN miss ratio converges to $\mu^*(\tau_0)$. It can then be shown (by analogy with (2.5)) that if we hold fixed a "normalized capacity" β_0 (which is the ratio of the expected number of pages in main memory under VMIN, divided by the number n of pages), then the expected VMIN miss ratio converges (as $n \to \infty$) to $\mu^*(\beta^{**-1}(\beta_0))$.

Let us apply these results to Zipf's law example of Table 1 (where the cumulative probability distribution function F is given by $F(x) = x^{0.5}$, and where the "normalized capacity" is 0.6). As we saw, the limit (as $n \to \infty$) of the expected LRU and expected working-set miss ratio is $\mu^*(\beta^{*-1}(0.6)) = 0.2902$. The limit (as $n \to \infty$) of the expected A_0 miss ratio is $1 - F(0.6) = 0.2254$, and the limit of the expected VMIN miss ratio is $\mu^*(\beta^{**-1}(0.6)) = 0.0973$. Thus, under a Zipf's law distribution with skewness 0.5, if the number of pages is large and if 60% of the pages can fit in main memory, then the expected LRU and working-set miss ratios are each approximately 0.2902, the expected A_0 miss ratio is approximately 0.2254, and the expected VMIN miss ratio is approximately 0.0973.

Since the formulas for $\beta^*(\tau_0)$ and $\mu^*(\tau_0)$ in the arithmetic case can be written in closed form without integrals (formulas (3.4)), it is amusing to look at this case a little closer. As before, let MISS(β_0) be the limit (as $n \to \infty$) of the expected LRU miss ratio (and of the expected working-set miss ratio), let $A_0(\beta_0)$ be the limit of the expected A_0 miss ratio, and let VMIN(β_0) be the limit of the expected VMIN miss ratio. It can be shown by using (3.4) that as $\beta_0 \to 1$ (which corresponds to $\tau_0 \to \infty$),

$$\begin{aligned} \mathrm{MISS}(\beta_0) &\to 2(1 - \beta_0)^2, \\ A_0(\beta_0) &\to (1 - \beta_0)^2, \\ \mathrm{VMIN}(\beta_0) &\to (\tfrac{1}{2})(1 - \beta_0)^2. \end{aligned}$$

In particular, under an arithmetic probability distribution, if the number of pages is large and if a very high percentage of the pages can fit in main memory, then the expected LRU and expected working-set miss ratios are approximately double the expected A_0 miss ratio, which in turn is approximately double the expected VMIN miss ratio.

5. Independence of LRU Miss Ratio from Page Size

From our results, it follows that asymptotically, LRU miss ratio in the independent reference model depends on two factors: The cumulative probability distribution function F and the fraction β_0 of pages that can fit in main memory. We will show in this section that this implies that LRU miss ratio in the independent reference model is approximately independent of page size if the size of main memory (in bytes) is held fixed, provided pages are blocked together in order of their probabilities of reference.

We will deal in this section with concave cumulative probability distribution functions F. Let $\{p_1, ..., p_n\}$ be the probability distribution determined by F and n (so $p_1 \geqslant \cdots \geqslant p_n$, by concavity of F). Let LMISS be the LRU miss ratio in the independent reference model with probability distribution $\{p_1, ..., p_n\}$ and with capacity CAP pages. Under our previous terminology, LMISS is LRU(n, (CAP/n)). We will now consider the effect of moving all data in blocks, each of which contains B pages. Our rule for block formation is to combine 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. Assume for convenience that B divides both n, the number of pages, and CAP, the capacity.

Since $p_1 \geqslant \cdots \geqslant p_n$, the blocked case amounts to dealing with the independent reference model with probability distribution $\{u_1, ..., u_{n/B}\}$, where $u_1 = p_1 + \cdots + p_B$, $u_2 = p_{B+1} + \cdots + p_{2B}$, etc. But then $\{u_1, ..., u_{n/B}\}$ is simply the probability distribution determined by F and n/B. Let LMISS* be the LRU miss ratio in the independent reference model with probability distribution $\{u_1, ..., u_{n/B}\}$ and capacity 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). Under our previous terminology, LMISS* is LRU((n/B), (CAP/n)). If n (and n/B) are sufficiently large, then LMISS and LMISS* are each as close as desired to the asymptotic LRU miss ratio with cumulative probability distribution function F and with $\beta_0 = \text{CAP}/n$. Hence, LMISS $\approx$ LMISS*, that is, the LRU miss ratio in the blocked and unblocked cases are approximately the same.

Now let WMISS and WMISS* be the expected working-set miss ratios in the unblocked and blocked cases with expected working-set sizes CAP pages and CAP/B blocks, respectively. Under our previous terminology, WMISS = WORK(n, (CAP/n)) and WMISS* = WORK((n/B), (CAP/n)). By the same argument as above, we find that WMISS $\approx$ WMISS*. Indeed, it was proven in [7] that if $C = \text{CAP}/B$ is the expected working-set size in blocks, then

$$|\,\text{WMISS} - \text{WMISS*}\,| < (2/C) + (33/C^2).$$

In [7] an argument is presented that if LRU miss ratio is insensitive to page size in the independent reference model, then the same insensitivity is to be expected in more realistic models of page reference strings, in which page references have an "independent" component and a "sequential" component. (In addition, some justification is given there to the assumption that pages are blocked together in order of their reference probabilities.)

Hence, the results in this section hold under less restrictive assumptions than those of the independent reference model.

6. Extension to a Model with Locality

Several papers [3, 16] have dealt with the difference between LRU memory management and working-set memory management. Under multiprogramming, if each program is managed by LRU, then this is a fixed partition of main memory; if each program is managed by a working-set memory management policy, then this is a corresponding variable partition of main memory. By the results in this paper, in the independent reference model the miss ratios in the two cases get arbitrarily close as the number of pages gets large. We will now generalize this result to a model (due to Easton [5]) with "locality of reference" [4].

Easton's model is given by a first-order Markov chain. If there are n pages, then there are $n+1$ parameters, $r, p_1, \ldots, p_n$, all lying between 0 and 1. Assume that page i was referenced at time t. At time $t+1$, the process goes into "rereference" (or "sequential") mode with probability r, and page i is rereferenced. With probability $(1-r)$, the process goes into "random" mode, and page j is referenced with probability $(1-r)p_j$, for $1 \leqslant j \leqslant n$ (including the case $j=i$). 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

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

Easton found [5] that with an appropriate choice of parameters, if capacities are reasonably large, then his model gives a good fit to the LRU miss ratio curve of the database portion of the IBM Advanced Administration System (A.A.S.) [19], a large database system. Intuitively, this model "works" because if the page size is large enough, then locality can be approximately captured by rereferences to the same page.

It is not hard to see that the expected LRU miss ratio in Easton's model is $(1-r)$ times the expected LRU miss ratio in the underlying independent reference model, where r is the probability of going into "rereference" mode. This is because rereferences do not change the LRU stack. Also, the expected working-set miss ratio in Easton's model gets arbitrarily close to $(1-r)$ times the expected working-set miss ratio in the underlying independent reference model, as the number of pages gets large. This latter result follows by obtaining formulas for Easton's model analogous to $\mu(n, \tau_0)$ and $\beta(n, \tau_0)$ of (2.8), and expressing the limits as integrals. Then the limit (as $n \to \infty$) of the expected working-set miss ratio in Easton's model turns out to be $(1-r)$ $\mathrm{MISS}(\beta_0)$, where $\mathrm{MISS}(\beta_0)$ is as before.

So the asymptotic equivalence of the LRU and working-set miss ratios in the independent reference model carries over to Easton's model.

We conclude this section with another application of our results. Assume that we have a database (such as that of A.A.S., mentioned earlier in this section) for which Easton's model gives a good fit to the LRU miss ratio curve. Assume that there is reason to believe that as the database grows (that is, as the number of pages increases, with the page size remaining fixed), it happens that

(1) the shape of the cumulative probability distribution function remains approximately the same, and

(2) the appropriate Easton's model again gives a good fit to the LRU miss ratio curve, with the parameter r of Easton's model remaining approximately the same.

As for assumption (1), we will assume in particular that

(1′) the cumulative probability distribution function is based on Zipf's law with skewness $\theta \approx 0.86$, which corresponds to the "80/20 law" (see Section 3).

As for assumption (2), it is not unreasonable to assume that the parameter r remains about the same, since r is roughly a measure of sequentiality within a typical page, and hence should remain more or less constant if the applications and database organization do not change very much.

An interesting question is how the size of main memory must vary in order to maintain the same miss ratio. For example, say that when the database contains n pages of k bytes each, and when the capacity is CAP pages, then the LRU miss ratio turns out to be M. When, in a number of years, the database has doubled in size, that is, when it contains $2n$ pages of k bytes each, how big should the capacity be so that the LRU miss ratio remains approximately M? It follows immediately from our results that the new capacity should be $2 \cdot \text{CAP}$, since then the "normalized capacity" $(2 \cdot \text{CAP})/(2 \cdot n)$ remains the same, and since the miss ratio is essentially determined by the normalized capacity. In other words, if the size of the database and the size of main memory (in bytes) each double, and if assumptions (1′) and (2) both hold, then the LRU miss ratio remains approximately the same.

7. Further Extensions

There are various further directions in which the results of Section 2 can be extended. We remark briefly on three such generalizations without giving details.

(1) Page sizes need not be all equal. This corresponds to dividing the x axis in Fig. 1 into n intervals whose lengths need not be equal (however, as $n \to \infty$, the maximum interval length must go to 0). All asymptotic results go through easily.

(2) The LRU page replacement algorithm may be generalized to allow page fixing, in which certain key pages always remain in main memory. This corresponds to allowing cumulative probability distribution functions F for which $F(1) < 1$. Again, all asymptotic results go through, with certain natural modifications, such as changing the limits of integration.

(3) Under our definition of a cumulative probability distribution function F, the derivative F' is continuous and finite (except possibly $F'(0) = \infty$.) We can drop all finiteness restrictions on F': Thus, F' is allowed to be infinite at any number of points. The essential reason that all of our results still hold is that every monotonic function (in our case, F) has a finite derivative except on a set of measure 0 [2, p. 134].

8. Proofs

In this section, we prove that if the cumulative probability distribution function F and the fraction β_0 of pages that can fit in main memory are held fixed (β_0 is the capacity divided by the total number of pages), then the expected LRU miss ratio converges, as the number of pages goes to infinity. In addition, we show that the expected working-set miss ratio, when β_0 is the expected working-set size divided by the number of pages, also converges to the same limit, as the number of pages goes to infinity. Their common limit is shown to be $\mu^*(\beta^{*-1}(\beta_0))$, where

$$\beta^*(\tau_0) = 1 - \int_0^1 e^{-\tau_0 F'(x)}\, dx, \qquad 0 \leqslant \tau_0 \leqslant \infty,$$

$$\mu^*(\tau_0) = \int_0^1 F'(x)\, e^{-\tau_0 F'(x)}\, dx, \qquad 0 \leqslant \tau_0 \leqslant \infty.$$

Let F be a cumulative probability distribution function, and $\tau_0 \geqslant 0$ a real number. Let $\{p_1^{(n)},..., p_n^{(n)}\}$ be the probability distribution determined by F and n. As in (2.8), define

$$\mu(n, \tau_0) = \sum_{i=1}^{n} p_i^{(n)}(1 - p_i^{(n)})^{\tau_0 n},$$

$$\beta(n, \tau_0) = 1 - (1/n) \sum_{i=1}^{n} (1 - p_i^{(n)})^{\tau_0 n}.$$

As we remarked earlier, if $T = \tau_0 n$ is an integer and T is the window size, then $\mu(n, \tau_0)$ is the expected working-set miss ratio, and $\beta(n, \tau_0)$ is the ratio of the expected working-set size divided by the number of pages.

We begin our proofs of the results in this paper by showing that if F and τ_0 (as opposed to β_0) are held fixed, then $\beta(n, \tau_0)$ and $\mu(n, \tau_0)$ each converge, as the number n of pages gets large. The limit of $\beta(n, \tau_0)$ is

$$\beta^*(\tau_0) = 1 - \int_0^1 e^{-\tau_0 F'(x)}\, dx,$$

and the limit of $\mu(n, \tau_0)$ is

$$\mu^*(\tau_0) = \int_0^1 F'(x)\, e^{-\tau_0 F'(x)}\, dx.$$

THEOREM 1.

$$\beta(n, \tau_0) \to \beta^*(\tau_0), \qquad \text{as} \quad n \to \infty, \tag{8.1}$$

$$\mu(n, \tau_0) \to \mu^*(\tau_0), \qquad \text{as} \quad n \to \infty. \tag{8.2}$$

Proof. We first prove (8.1). Statement (8.1) is trivially equivalent to

$$(1/n) \sum_{i=1}^{n} (1 - p_i^{(n)})^{\tau_0 n} \to \int_0^1 e^{-\tau_0 F'(x)}\, dx, \qquad \text{as} \quad n \to \infty. \tag{8.3}$$

Due to technical problems which arise in case $F'(0) = \infty$, it is convenient to show that for each δ, with $0 < \delta < 1$,

$$(1/n) \sum_{i=\lfloor \delta n \rfloor + 1}^{n} (1 - p_i^{(n)})^{\tau_0 n} \to \int_\delta^1 e^{-\tau_0 F'(x)}\, dx, \qquad \text{as} \quad n \to \infty. \tag{8.4}$$

Then (8.4) is sufficient to prove (8.3), since it is easy to verify that the left-hand side of (8.4) differs from the left-hand side of (8.3) by at most δ (since $(1 - p_i)^{\tau_0 n} \leqslant 1$), and that the right-hand side of (8.4) differs from the right-hand side of (8.3) by at most δ (since $F' \geqslant 0$, and hence $e^{-\tau_0 F'(x)} \leqslant 1$). Then (8.3) follows from (8.4) by letting δ go to 0. We leave the (straightforward) details to the reader.

We will now prove (8.4). For ease in notation, write p_i for $p_i^{(n)}$, and write d for $\lfloor \delta n \rfloor + 1$. By the Mean Value Theorem,

$$p_i = F'(\zeta_i)/n, \tag{8.5}$$

for some ζ_i with $((i - 1)/n) \leqslant \zeta_i \leqslant (i/n)$.

We will show in Appendix 3 that for each closed bounded set B,

$$(1 - (b/n)^{\tau_0 n} \to e^{-\tau_0 b} \qquad \text{as } n \to \infty, \text{ uniformly over all } b \text{ in } B. \tag{8.6}$$

Let $B = \{F'(x): (\delta/2) \leqslant x \leqslant 1\}$. By continuity of F', the set B is closed and bounded. It is clear that there is N_0 so large (in fact $N_0 = 2/\delta$ is adequate) that if $n > N_0$, then $\zeta_i \geqslant \delta/2$ for each $i > d$, and hence $F'(\zeta_i) \in B$. Pick $\epsilon > 0$. By (8.6), find $N_1 > N_0$ so large that for each $n > N_1$ and each b in B, we have

$$|(1 - (b/n))^{\tau_0 n} - e^{-\tau_0 b}| < \epsilon.$$

In particular, if we define

$$\mathrm{ERR}(i, n) = (1 - (F'(\zeta_i)/n))^{\tau_0 n} - e^{-\tau_0 F'(\zeta_i)} \tag{8.7}$$

then

$$|\mathrm{ERR}(i, n)| < \epsilon.$$

Then

$$\begin{aligned}
&\left| (1/n) \sum_{i=d}^{n} (1 - p_i)^{\tau_0 n} - (1/n) \sum_{i=d}^{n} e^{-\tau_0 F'(\zeta_i)} \right| \\
&\quad = \left| (1/n) \sum_{i=d}^{n} (1 - (F'(\zeta_i)/n))^{\tau_0 n} - (1/n) \sum_{i=d}^{n} e^{-\tau_0 F'(\zeta_i)} \right|, \qquad \text{by (8.5)}, \\
&\quad = \left| (1/n) \sum_{i=d}^{n} \mathrm{ERR}(i, n) \right|, \qquad \text{by (8.7)}, \\
&\quad \leqslant (1/n) \sum_{i=d}^{n} |\mathrm{ERR}(i, n)|, \qquad \text{by the triangle inequality}, \\
&\quad < (1/n)(n)\, \epsilon, \\
&\quad = \epsilon.
\end{aligned} \tag{8.8}$$

Now by definition of the Riemann integral,

$$(1/n)\sum_{i=d}^{n} e^{-\tau_0 F'(\zeta_i)} = (1/n)\sum_{i=\lfloor\delta n\rfloor+1}^{n} e^{-\tau_0 F'(\zeta_i)}$$

is an approximation to the Riemann integral $\int_\delta^1 e^{-\tau_0 F'(x)}\,dx$, where the interval $[\delta, 1]$ is broken into subintervals of length $1/n$. So we can find $N_2 > N_1$ so large that for each $n > N_2$,

$$\left|(1/n)\sum_{i=d}^{n} e^{-\tau_0 F'(\zeta_i)} - \int_\delta^1 e^{-\tau_0 F'(x)}\,dx\right| < \epsilon. \tag{8.9}$$

Hence for all $n > N_2$,

$$\begin{aligned}
&\left|(1/n)\sum_{i=d}^{n}(1-p_i)^{\tau_0 n} - \int_\delta^1 e^{-\tau_0 F'(x)}\,dx\right| \\
&\quad\leqslant \left|(1/n)\sum_{i=d}^{n}(1-p_i)^{\tau_0 n} - (1/n)\sum_{i=d}^{n} e^{-\tau_0 F'(\zeta_i)}\right| + \left|(1/n)\sum_{i=d}^{n} e^{-\tau_0 F'(\zeta_i)}\right. \\
&\quad\quad \left. - \int_\delta^1 e^{-\tau_0 F'(x)}\,dx\right|, \qquad \text{by the triangle inequality,} \\
&\quad< \epsilon + \epsilon, \qquad \text{by (8.8) and (8.9),} \\
&\quad= 2\epsilon,
\end{aligned}$$

Then (8.4) follows, as desired.

We will now prove (8.2). Statement (8.2) says

$$\sum_{i=1}^{n} p_i^{(n)}(1-p_i^{(n)})^{\tau_0 n} \to \int_0^1 F'(x)\, e^{-\tau_0 F'(x)}\,dx, \qquad \text{as} \quad n \to \infty. \tag{8.10}$$

Let δ be arbitrary, $0 < \delta < 1$. We will prove that

$$\sum_{i=\lfloor\delta n\rfloor+1}^{n} p_i^{(n)}(1-p_i^{(n)})^{\tau_0 n} \to \int_\delta^1 F'(x)\, e^{-\tau_0 F'(x)}\,dx, \qquad \text{as} \quad n \to \infty. \tag{8.11}$$

Then (8.11) implies (8.10). For, the left-hand side of (8.11) differs from the left-hand side of (8.10) by at most $\sum_{i=1}^{\lfloor\delta n\rfloor} p_i^{(n)}$, which is bounded by $F(\delta)$, which goes to 0 as δ goes to 0. And, the right-hand side of (8.11) differs from the right-hand side of (8.10) by at most δ/e, since the maximum possible value of $F'(x)\, e^{-F'(x)}$ (or of ze^{-z} for arbitrary z) is $1/e$, by elementary calculus.

We will now prove (8.11). Pick $\epsilon > 0$. Write $d = \lfloor\delta n\rfloor + 1$, and write p_i for $p_i^{(n)}$. Find ζ_i as in (8.5), for each i.

Let $C = \sup\{F'(x): (\delta/2) \leqslant x \leqslant 1\}$. As before, we can choose N_0 so large that if

$n > N_0$, then $\zeta_i \geqslant \delta/2$ for $i > d$, and hence $F'(\zeta_i) \leqslant C$. If ERR(i, n) is as before, then we can again find $N_1 > N_0$ so large that for all $n > N_1$,

$$|\operatorname{ERR}(i, n)| < \epsilon/C.$$

So,

$$\begin{aligned}
&\left| \sum_{i=d}^{n} p_i(1 - p_i)^{\tau_0 n} - (1/n) \sum_{i=d}^{n} F'(\zeta_1)\, e^{-\tau_0 F'(\zeta_i)} \right| \\
&\quad = \left| (1/n) \sum_{i=d}^{n} F_i'(\zeta_i)(1 - (F'(\zeta_i)/n))^{\tau_0 n} - (1/n) \sum_{i=d}^{n} F'(\zeta_i)\, e^{-\tau_0 F'(\zeta_i)} \right|, \qquad \text{by (8.5)}, \\
&\quad = \left| (1/n) \sum_{i=d}^{n} F_i'(\zeta_i) \operatorname{ERR}(i, n) \right|, \qquad (8.12) \\
&\quad \leqslant (1/n) \sum_{i=d}^{n} C\, |\operatorname{ERR}(i, n)|, \qquad \text{by the triangle inequality}, \\
&\quad < (1/n)(n)(C)(\epsilon/C), \\
&\quad = \epsilon.
\end{aligned}$$

As before, we can find $N_2 > N_1$ so large that for all $n > N_2$,

$$\left| (1/n) \sum_{i=d}^{n} F'(\zeta_i)\, e^{-\tau_0 F'(\zeta_i)} - \int_{\delta}^{1} F'(x)\, e^{-\tau_0 F'(x)}\, dx \right| < \epsilon. \tag{8.13}$$

From (8.12) and (8.13), we find that for all $n > N_2$,

$$\left| \sum_{i=d}^{n} p_i(1 - p_i)^{\tau_0 n} - \int_{\delta}^{1} F'(x)\, e^{-\tau_0 F'(x)}\, dx \right| < \epsilon + \epsilon = 2\epsilon.$$

Then (8.11) follows, as desired. ∎

In Theorem 1, we held τ_0 fixed. However, we are really interested in holding β_0 fixed. As before, define WORK(n, β_0) to be the expected working-set miss ratio with expected working-set size $\lfloor \beta_0 n \rfloor$, over the probability distribution $\{p_1^{(n)}, \ldots, p_n^{(n)}\}$ determined by F and n. We can now prove that WORK(n, β_0) converges to MISS$(\beta_0) = \mu^*(\beta^{*-1}(\beta_0))$, as the number n of pages gets large.

THEOREM 2. *WORK$(n, \beta_0) \to$ MISS(β_0), as $n \to \infty$.*

Proof. For each positive integer n, find τ_n such that

$$\beta(n, \tau_n) = \lfloor \beta_0 n \rfloor / n.$$

This is possible, because $\beta(n, \tau)$ takes on all values between 0 and 1 as τ ranges between 0 and ∞. By definition,

$$\mathrm{WORK}(n, \beta_0) = \mu(n, \tau_n). \tag{8.14}$$

Find τ_0 such that $\beta^*(\tau_0) = \mathrm{MISS}(\beta_0)$. We will show that

$$\tau_n \to \tau_0, \qquad \text{as} \quad n \to \infty. \tag{8.15}$$

The fact that τ_n converges is implicit in the proof in [7]. We will now demonstrate that (8.15) is sufficient to prove the theorem.

Pick $\epsilon > 0$. By continuity of μ^*, we can find $\delta > 0$ so small that

$$\mu^*(\tau_0 - \delta) < \mu^*(\tau_0) + \epsilon. \tag{8.16}$$

By Theorem 1, with $\tau_0 - \delta$ playing the role of τ_0, we know that

$$\mu(n, \tau_0 - \delta) \to \mu^*(\tau_0 - \delta), \qquad \text{as} \quad n \to \infty.$$

Hence, we can find N so large that if $n > N$, then

$$\mu(n, \tau_0 - \delta) < \mu^*(\tau_0 - \delta) + \epsilon. \tag{8.17}$$

By (8.15), we can also assume that N is so large that if $n > N$, then

$$\tau_n > \tau_0 - \delta. \tag{8.18}$$

Assume that $n > N$. Since μ is monotone decreasing in its second argument, it follows from (8.18) that

$$\begin{aligned} \mu(n, \tau_n) &\leqslant \mu(n, \tau_0 - \delta) \\ &< \mu^*(\tau_0 - \delta) + \epsilon, \qquad \text{by} \quad (8.17), \\ &< \mu^*(\tau_0) + 2\epsilon, \qquad \text{by} \quad (8.16). \end{aligned} \tag{8.19}$$

We can substitute $\mathrm{WORK}(n, \beta_0)$ for $\mu(n, \tau_n)$ in (8.19) by (8.14), and we can substitute $\mathrm{MISS}(\beta_0)$ for $\beta^*(\tau_0)$ by definition of τ_0. We then obtain

$$\mathrm{WORK}(n, \beta_0) < \mathrm{MISS}(\beta_0) + 2\epsilon. \tag{8.20}$$

By a symmetric argument, we find that if n is sufficiently large, then

$$\mathrm{WORK}(n, \beta_0) > \mathrm{MISS}(\beta_0) - 2\epsilon. \tag{8.21}$$

The theorem follows from (8.20) and (8.21). It remains to prove (8.15).

Assume (8.15) fails. Then for some $\epsilon > 0$, either

$$\tau_n < \tau_0 - \epsilon, \qquad \text{for infinitely many } n, \tag{8.22}$$

or

$$\tau_n > \tau_0 + \epsilon, \qquad \text{for infinitely many } n. \tag{8.23}$$

Assume that (8.22) holds; the proof if (8.23) holds is symmetric. Let $\tau' = \tau_0 - (\epsilon/2)$, and $\beta' = \beta^*(\tau')$. By monotonicity of β^*,

$$\beta' = \beta^*(\tau') < \beta^*(\tau_0) = \beta_0 .$$

Let $\theta = \beta_0 - \beta' > 0$. By Theorem 1,

$$\beta(n, \tau') \to \beta^*(\tau') = \beta', \qquad \text{as} \quad n \to \infty.$$

So for n sufficiently large,

$$\beta(n, \tau') < \beta' + (\theta/2) = \beta_0 - (\theta/2). \tag{8.24}$$

By (8.22) and (8.24), there is some integer N such that

$$\tau_N < \tau_0 - \epsilon, \tag{8.25}$$

$$\beta(N, \tau') < \beta_0 - (\theta/2). \tag{8.26}$$

We can also assume that N is so large that

$$\lfloor \beta_0 N \rfloor / N > \beta_0 - (\theta/2). \tag{8.27}$$

Then

$$\begin{aligned}
\lfloor \beta_0 N \rfloor / N &= \beta(N, \tau_N), && \text{by definition of } \tau_N , \\
&\leqslant \beta(N, \tau_0 - \epsilon), && \text{by (8.25) and monotonicity,} \\
&\leqslant \beta(N, \tau'), && \text{by monotonicity and definition of } \tau', \\
&< \beta_0 - (\theta/2), && \text{by (8.26),} \\
&< \lfloor \beta_0 N \rfloor / N, && \text{by (8.27).}
\end{aligned}$$

This is a contradiction. ∎

As before, define $\mathrm{LRU}(n, \beta_0)$ to be the LRU miss ratio with capacity $\lfloor \beta_0 n \rfloor$ over the probability distribution $\{p_1^{(n)}, \ldots, p_n^{(n)}\}$ determined by F and n. We will now prove that $\mathrm{LRU}(n, \beta_0)$ converges to $\mathrm{MISS}(\beta_0) = \mu^*(\beta^{*-1}(\beta_0))$, as the number of pages gets large. We begin with a lemma, which is due to Easton [5].

LEMMA 3 (Easton). *Under working-set memory management in the independent reference model with n pages, the variance in the size of the working set is bounded above by n/4.*

Proof. Let T be the window size, and consider the process in which there are T independent page references.

Let X be a random variable which is equal to the number of distinct pages referenced over the course of the T references. Thus, X is the working-set size. We went to show that the variance $V(X)$ of X is bounded above by $n/4$.

Let X_i, $1 \leqslant i \leqslant n$, be a random variable which is 1 if page i is in the working-set (that is, if page i appears in the course of T references), and 0 otherwise. So $X = X_1 + \cdots + X_n$. Let A_i be the event that page i occurs. So, $E(X_i)$, the expected value of X_i, equals $\Pr[A_i]$, the probability of event A_i. Then the variance $V(X)$ in the size of the working set is

$$V(X_1 + \cdots + X_n) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j), \tag{8.28}$$

where Cov is the covariance [8, p. 215]. Now

$$V(X_i) \leqslant \tfrac{1}{4}, \tag{8.29}$$

since X_i is 0-1 valued. (If $X_i = 1$ with probability p, then $V(X_i) = p(1-p) \leqslant \frac{1}{4}$ by elementary calculus); and

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)\,E(X_j) \\ &= \Pr[A_i \wedge A_j] - \Pr[A_i]\Pr[A_j] \\ &= \Pr[A_i]\Pr[A_j \mid A_i] - \Pr[A_i]\Pr[A_j] \\ &= \Pr[A_i](\Pr[A_j \mid A_i] - \Pr[A_j]) \\ &\leqslant 0, \end{aligned} \tag{8.30}$$

since if page i appears, then this lessens the probability that page j appears.

So from (8.28), (8.29), and (8.30), we find that $V(X_1 + \cdots + X_n) \leqslant n/4$, as desired. ∎

Remark. If $p_1 = \cdots = p_n = 1/n$, and $T = \lfloor (\ln 2)n \rfloor \approx 0.693n$, then the variance in the size of the working set can be shown to be arbitrarily close to $n/4$ for n sufficiently large.

THEOREM 4. *$\mathrm{LRU}(n, \beta_0) \to \mathrm{MISS}(\beta_0)$, as $n \to \infty$.*

Proof. Pick $\epsilon > 0$. Find τ_0 such that $\beta^*(\tau_0) = \beta_0$, and let $\mu_0 = \mu^*(\tau_0) = \mathrm{MISS}(\beta_0)$. Find $\tau_{\mathrm{big}} > \tau_0$ so close to τ_0 that $\mu_0 - \mu^*(\tau_{\mathrm{big}}) < \epsilon$. Write $\mu_{\mathrm{big}} = \mu^*(\tau_{\mathrm{big}})$, and $\beta_{\mathrm{big}} = \beta^*(\tau_{\mathrm{big}})$. So

$$\mu_0 - \mu_{\mathrm{big}} < \epsilon. \tag{8.31}$$

(Warning: μ_{big} is *smaller* than μ_0, because μ^* is a monotone *decreasing* function of τ.) Let $\theta = \beta_{\mathrm{big}} - \beta_0 > 0$. By Theorem 1, if n is sufficiently large, then

$$|\beta(n, \tau_{\mathrm{big}}) - \beta_{\mathrm{big}}| < \theta/2 \tag{8.32}$$

and

$$|\mu(n, \tau_{\mathrm{big}}) - \mu_{\mathrm{big}}| < \epsilon. \tag{8.33}$$

Assume also that n is sufficiently large that both

$$n > 1/(\theta^2 \epsilon) \tag{8.34}$$

and also that if $\{p_1, \ldots, p_n\}$ is the probability distribution determined by F and n, then

$$\max\{p_1, \ldots, p_n\} < \epsilon. \tag{8.35}$$

Consider the following process, which we will call Process (*).

(*) There are n pages, with reference probabilities $\{p_1, ..., p_n\}$ determined by F and n, and a sequence of independent references.

Let $T_{\text{big}} = \lceil \tau_{\text{big}} n \rceil$, and let CAP $= \lfloor \beta_0 n \rfloor$. Define events A, B, and C as

A is the event that during the first T_{big} references in Process (*), at least CAP distinct pages appeared.

B is the event that the $(T_{\text{big}} + 1)$st reference already appeared among the first T_{big}, in Process (*).

C is the event that just after there have been CAP distinct pages referenced for the first time, the next reference is to one of the CAP pages that has already appeared, in Process (*).

Then $\Pr[C]$ is the LRU hit ratio with capacity CAP, that is,

$$\Pr[C] = 1 - \mathrm{LRU}(n, \beta_0). \tag{8.36}$$

Also, $\Pr[B]$ is the working-set hit ratio with window-size T_{big}. The following estimate is helpful.

$$\begin{aligned}
\Pr[B] &= 1 - \sum p_i(1-p_i)^{T_{\text{big}}}, \\
&= 1 - \sum p_i(1-p_i)^{\lfloor \tau_{\text{big}} n \rfloor}, \\
&\leqslant 1 - \sum p_i(1-p_i)^{\tau_{\text{big}} n} + \epsilon, \qquad \text{by (8.35) and a simple argument,} \\
&= 1 - \mu(n, \tau_{\text{big}}) + \epsilon, \\
&\leqslant 1 - \mu_{\text{big}} + 2\epsilon, \qquad \text{by (8.33)}, \\
&\leqslant 1 - \mu_0 + 3\epsilon, \qquad \text{by (8.31)}.
\end{aligned} \tag{8.37}$$

We will now show that A is a likely event. Consider again Process (*) defined earlier, and let X be a random variable whose value is the number of distinct references among the first T_{big}. The expected value $E(X)$ is

$$\begin{aligned}
E(X) &= n - \sum (1-p_i)^{T_{\text{big}}}, \qquad \text{by (1.3)}, \\
&= n - \sum (1-p_i)^{\lceil \tau_{\text{big}} n \rceil}, \\
&\geqslant n - \sum (1-p_i)^{\tau_{\text{big}} n}, \\
&= n\beta(n, \tau_{\text{big}}), \\
&> n(\beta_{\text{big}} - (\theta/2)), \qquad \text{by (8.32)}.
\end{aligned} \tag{8.38}$$

Hence,

$$\begin{aligned}
E(X) - \mathrm{CAP} &= E(X) - \lfloor \beta_0 n \rfloor, \\
&> n(\beta_{\text{big}} - (\theta/2)) - \lfloor \beta_0 n \rfloor, \qquad \text{by (8.38)}, \\
&\geqslant n(\beta_{\text{big}} - (\theta/2)) - \beta_0 n, \\
&= n\theta/2, \qquad \text{since } \theta = \beta_{\text{big}} - \beta_0 .
\end{aligned} \tag{8.39}$$

Now for the event $\sim A$ to occur, X must take on a value smaller than CAP $= \lfloor \beta_0 n \rfloor$. Therefore, by (8.39) it follows that for event $\sim A$ to occur, X must differ from $E(X)$ by at least $n\,\theta/2$. Hence,

$$\begin{aligned} \Pr[\sim A] &\leqslant \Pr[|X - E(X)| \geqslant n\,\theta/2], \\ &\leqslant 4V(X)/(n^2\theta^2), && \text{by Chebyshev's inequality [8, p. 219],} \\ & && \text{where } V(X) \text{ is the variance of } X, \\ &\leqslant 1/(n\theta^2), && \text{since } V(X) \leqslant n/4 \text{ by Lemma 3,} \\ &< \epsilon, && \text{by (8.34).} \end{aligned} \tag{8.40}$$

We now interrelate $\Pr[A]$, $\Pr[B]$, and $\Pr[C]$ as follows.

$$\begin{aligned} \Pr[C] &= \Pr[C \wedge A] + \Pr[C \wedge \sim A], \\ &\leqslant \Pr[C \wedge A] + \Pr[\sim A], \\ &\leqslant \Pr[B \wedge A] + \Pr[\sim A], \qquad \text{since} \quad \Pr[C \mid A] \leqslant \Pr[B \mid A], \\ &\leqslant \Pr[B] + \Pr[\sim A]. \end{aligned} \tag{8.41}$$

Putting together what we have proved, we obtain

$$\begin{aligned} 1 - \mathrm{LRU}(n, \beta_0) &= \Pr[C], && \text{by (8.36),} \\ &\leqslant \Pr[B] + \Pr[\sim A], && \text{by (8.41),} \\ &\leqslant 1 - \mu_0 + 3\epsilon + \epsilon, && \text{by (8.37) and (8.40).} \end{aligned}$$

Hence, for n sufficiently large,

$$\begin{aligned} \mathrm{LRU}(n, \beta_0) &\geqslant \mu_0 - 4\epsilon \\ &= \mathrm{MISS}(\beta_0) - 4\epsilon. \end{aligned} \tag{8.42}$$

By a very similar argument, we can show that

$$\mathrm{LRU}(n, \beta_0) \leqslant \mathrm{MISS}(\beta_0) + 4\epsilon. \tag{8.43}$$

The only essential modification of the proof is to change each occurrence of T_{big} in the definitions of events A, B, and C to T_{little}, where $T_{\text{little}} = \lfloor \tau_{\text{little}} n \rfloor$ and $\tau_{\text{little}} < \tau_0$ is close to τ_0. Then derivation (8.41) is replaced by

$$\begin{aligned} \Pr[B] &= \Pr[B \wedge A] + \Pr[B \wedge \sim A] \\ &\leqslant \Pr[A] + \Pr[B \wedge \sim A] \\ &\leqslant \Pr[A] + \Pr[C \wedge \sim A] \\ &\leqslant \Pr[A] + \Pr[C]. \end{aligned}$$

Inequality (8.43) then follows.

By (8.42) and (8.43), we see that for n sufficiently large,

$$\text{MISS}(\beta_0) - 4\epsilon \leqslant \text{LRU}(n, \beta_0) \leqslant \text{MISS}(\beta_0) + 4\epsilon.$$

The theorem follows. ∎

Of course, we have the following immediate corollary of Theorems 2 and 4.

THEOREM 5. $|\text{LRU}(n, \beta_0) - \text{WORK}(n, \beta_0)| \to 0$, *as* $n \to \infty$.

That is, the LRU miss ratio and the appropriate corresponding working-set miss ratio get arbitrarily close, as the number of pages gets large.

APPENDIX 1

We will show that if $0 \leqslant \theta < 1$, then the function $F: x \to x^{1-\theta}$ corresponds to the class of Zipf's law distributions with skewness θ, in the sense that if $\{z_1^{(n)},..., z_n^{(n)}\}$ is the Zipf's law distribution with skewness θ and with n pages, then for each β_0, with $0 \leqslant \beta_0 \leqslant 1$,

$$\sum_{i=1}^{\lfloor \beta_0 n \rfloor} z_i^{(n)} \to F(\beta_0), \qquad \text{as} \quad n \to \infty. \tag{A1.1}$$

For,

$$\sum_{i=1}^{\lfloor \beta_0 n \rfloor} z_i^{(n)} = \text{NUM/DENOM}, \tag{A1.2}$$

where

$$\text{NUM} = \sum_{i=1}^{\lfloor \beta_0 n \rfloor} i^{-\theta},$$

$$\text{DENOM} = \sum_{i=1}^{n} i^{-\theta}.$$

If we multiply NUM and DENOM each by $n^{1-\theta}$ and regroup, we find that

$$\text{NUM/DENOM} = \text{NUM1/DENOM1}, \tag{A1.3}$$

where

$$\text{NUM1} = \sum_{i=1}^{\lfloor \beta_0 n \rfloor} (i/n)^{-\theta}\,(1/n),$$

$$\text{DENOM1} = \sum_{i=1}^{n} (i/n)^{-\theta}\,(1/n).$$

Now NUM1 is an approximation to the Riemann integral $\int_0^{\beta_0} x^{-\theta}\,dx$, where the interval $[0, \beta_0]$ is broken into subintervals of length $1/n$. So,

$$\text{NUM1} \to \int_0^{\beta_0} x^{-\theta}\,dx, \qquad \text{as} \quad n \to \infty.$$

Similarly,

$$\text{DENOM1} \to \int_0^1 x^{-\theta}\,dx, \qquad \text{as} \quad n \to \infty.$$

If we evaluate these integrals, we find

$$\text{NUM1/DENOM1} \to \beta_0^{1-\theta}, \qquad \text{as} \quad n \to \infty. \tag{A1.4}$$

Hence

$$\begin{aligned} \sum_{i=1}^{\lfloor \beta_0 n \rfloor} z_i^{(n)} &= \text{NUM/DENOM}, && \text{by (A1.2)},\\ &= \text{NUM1/DENOM1}, && \text{by (A1.3)},\\ &\to \beta_0^{1-\theta}, && \text{as} \quad n \to \infty \quad \text{by (A1.4)}, \end{aligned}$$

The assumption that $0 \leqslant \theta < 1$ is used implicitly in the evaluation of the integrals, since NUM1 and DENOM1 are infinite for $\theta \geqslant 1$.

We remark that in addition to the previous proof, there is a simple intuitive reason why the function $F\colon x \to x^{1-\theta}$ approximately gives a Zipf's law distribution with skewness θ $(0 \leqslant \theta < 1)$. Let $\{p_1, \ldots, p_n\}$ be the probability distribution determined by F and n. Then

$$\begin{aligned} p_i &= F(i/n) - F((i-1)/n),\\ &\approx (1/n)F'(i/n), \qquad \text{by the Mean Value Theorem},\\ &= (1/n)(1-\theta)(i/n)^{-\theta}, \end{aligned}$$

which is proportional to $i^{-\theta}$ for fixed θ and n.

Appendix 2

In this appendix, we sketch the proof of a theorem which generalizes Theorems 2 and 4. Assume that for each n, we have a probability distribution $\{z_1^{(n)}, \ldots, z_n^{(n)}\}$ such that $z_1^{(n)} \geqslant z_2^{(n)} \geqslant \cdots \geqslant z_n^{(n)}$. Assume that F is a cumulative probability distribution function for which

$$\sum_{i=1}^{\lfloor \beta_0 n \rfloor} z_i^{(n)} \to F(\beta_0) \qquad \text{as} \quad n \to \infty,$$

for each β_0 with $0 \leqslant \beta_0 \leqslant 1$.

Define $L(n, \beta_0)$ to be the expected LRU miss ratio for probability distribution $\{z_1^{(n)}, \ldots, z_n^{(n)}\}$ when the capacity is $\lfloor \beta_0 n \rfloor$. Define $W(n, \beta_0)$ to be the expected working-set miss ratio for probability distribution $\{z_1^{(n)}, \ldots, z_n^{(n)}\}$ when the expected working-set size is $\lfloor \beta_0 n \rfloor$. Let $\text{MISS} = \text{MISS}_F$ be as defined earlier (2.5).

THEOREM 6.

$$L(n, \beta_0) \to \text{MISS}(\beta_0), \qquad \textit{as} \quad n \to \infty, \tag{A2.1}$$

$$W(n, \beta_0) \to \text{MISS}(\beta_0), \qquad \textit{as} \quad n \to \infty. \tag{A2.2}$$

This theorem, which we invoked in Section 3, is a generalization of Theorems 2 and 4. It is also convenient to generalize Theorem 1, as follows. By analogy with $\mu(n, \tau_0)$ and $\beta(n, \tau_0)$ in (2.8), define $\mu_1(n, \tau_0)$ and $\beta_1(n, \tau_0)$ as

$$\mu_1(n, \tau_0) = \sum_{i=1}^{n} z_i^{(n)}(1 - z_i^{(n)})^{\tau_0 n},$$

$$\beta_1(n, \tau_0) = 1 - (1/n) \sum_{i=1}^{n} (1 - z_i^{(n)})^{\tau_0 n}.$$

Thus, if $T = \tau_0 n$ is an integer, then $\mu_1(n, \tau_0)$ is the expected working-set miss ratio with window size T over probability distribution $\{z_1^{(n)},..., z_n^{(n)}\}$, and $\beta_1(n, \tau_0)$ is the expected working-set size divided by the number n of pages. In both cases, if T is not an integer, then these are interpolated values.

Define $\mu^* = \mu_F{}^*$ and $\beta^* = \beta_F{}^*$ as in (2.4).

Theorem 1 generalizes as

THEOREM 7.

$$\beta_1(n, \tau_0) \to \beta^*(\tau_0), \qquad as \quad n \to \infty, \tag{A2.3}$$

$$\mu_1(n, \tau_0) \to \mu^*(\tau_0), \qquad as \quad n \to \infty. \tag{A2.4}$$

Then Theorem 6 follows from Theorem 7 by almost the identical proof that Theorems 2 and 4 follow from Theorem 1. It remains to prove Theorem 7.

We will now state the key lemma by which the proof of Theorem 1 can be modified to prove Theorem 7. Since $z_1^{(n)} \geqslant z_2^{(n)} \geqslant \cdots \geqslant z_n^{(n)}$, it is possible to define a (smooth) concave cumulative probability distribution function F_n which "passes through" these points, in the sense that $F_n(i/n) - F_n((i-1)/n) = z_i^{(n)}$, for $1 \leqslant i \leqslant n$. Thus, $\{z_1^{(n)},...,z_n^{(n)}\}$ is the (unique) probability distribution determined by F_n and n.

LEMMA 8. *Assume that* $0 < \delta < 1$. *Then* $F_n' \to F'$ *as* $n \to \infty$, ***uniformly on*** $[\delta, 1 - \delta]$.

Remark. It follows in particular that $F_n' \to F'$ pointwise on (0, 1), that is, $F_n'(x) \to F'(x)$ for each x with $0 < x < 1$. However, $F_n'(0)$ need not converge to $F'(0)$, and $F_n'(1)$ need not converge to $F'(1)$.

Before proving Lemma 8 we will indicate how the proof of Theorem 1 can be modified, via Lemma 8, to prove Theorem 7.

Very much as (8.1) of Theorem 1 follows from (8.4) in the proof of Theorem 1, it also happens that (A2.3) of Theorem 7 follows from

$$(1/n) \sum_{i=\lfloor \delta n \rfloor + 1}^{n-\lceil \delta n \rceil} (1 - z_i^{(n)})^{\tau_0 n} \to \int_{\delta}^{1-\delta} e^{-\tau_0 F'(x)}\, dx, \qquad \text{as} \quad n \to \infty. \tag{A2.5}$$

Just as $p_i^{(n)} = F'(\zeta_i)/n$ in the proof of Theorem 1, similarly $z_i^{(n)} = F_n'(\zeta_i)/n = (1 + \theta(i, n))F'(\zeta_i)/n$, where by Lemma 8, if n is sufficiently large then $|\theta(i, n)|$ is arbitrarily small, uniformly in i $(1 \leqslant i \leqslant n)$.

The remaining details to modify the proof of Theorem 1 to prove (A2.5) are straightforward. Similarly (A2.4) follows, and Theorem 7 is proven.

We must now prove Lemma 8.

Proof of Lemma 8. It is not hard to see that F_n converges to F pointwise on [0, 1], as $n \to \infty$.

It follows fairly easily that since each F_n is concave, so is F. We will now show that F_n' converges pointwise to F' on (0, 1), as $n \to \infty$. Assume not. Let x_0 be a point, $0 < x_0 < 1$, such that $F_n'(x_0)$ does not converge to $F'(x_0)$, as $n \to \infty$. Then there exists $\epsilon > 0$ such that either

$$F_n'(x_0) > F'(x_0) + \epsilon, \qquad \text{for infinitely many } n, \tag{A2.6}$$

or

$$F_n'(x_0) < F'(x_0) - \epsilon, \qquad \text{for infinitely many } n. \tag{A2.7}$$

Assume that (A2.6) holds; the proof if (A2.7) holds is very similar. Let BAD $=$ $\{n: F_n'(x_0) > F'(x_0) + \epsilon\}$; by (A2.6), BAD is an infinite set. It is easy to see geometrically (see Fig. 2) that it is possible to find ϵ_1 with $0 < \epsilon_1 \leqslant \epsilon$, and $y_1 > F(x_0)$ so close to $F(x_0)$,

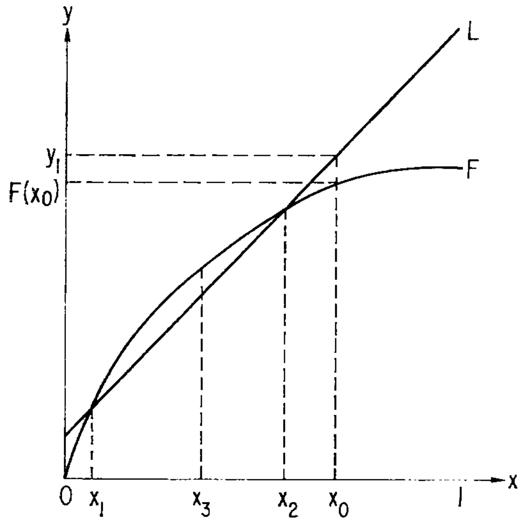

FIGURE 2

that the line L through (x_0, y_1) with slope $F'(x_0) + \epsilon_1$ intersects the graph of F twice. Let x_1 and x_2 be the x values of the intersection of F and L, and let x_3 be arbitrary, subject only to $x_1 < x_3 < x_2$. Find N so large that $F_n(x_0) < y_1$ for each $n > N$; this is possible since $F_n(x_0) \to F(x_0)$ as $n \to \infty$. Let BAD1 $= \{n > N: n \in \text{BAD}\}$. Now if $n \in$ BAD1, then $F_n'(x_0) > F'(x_0) + \epsilon \geqslant F'(x_0) + \epsilon_1 = L'(x_0)$. So if $n \in$ BAD1, then the tangent line to F_n at x_0 is steeper than L, and hence lies completely below L for $0 \leqslant x \leqslant x_0$. But F_n, being concave, lies below its tangent line. Hence F_n lies below L for $0 \leqslant x \leqslant x_0$, that is, $F_n(x) < L(x)$ for each n in BAD1 and each x between 0 and x_0. In particular,

$F_n(x_3) < L(x_3)$ for each n in BAD1. Hence, $F(x_3) - F_n(x_3) > F(x_3) - L(x_3) > 0$ for each n in BAD1. But then $F(x_3) - F_n(x_3)$ is bounded away from 0 by the positive quantity $F(x_3) - L(x_3)$, for each n in BAD1. This contradicts the fact that $F_n(x_3) \to F(x_3)$, as $n \to \infty$. So we have shown that F_n' converges to F' pointwise on (0, 1), and hence on $[\delta, 1 - \delta]$. To show that F_n' converges to F' uniformly on $[\delta, 1 - \delta]$, the following lemma is clearly sufficient, where G_n is F_n' and G is F'.

LEMMA 9. *Assume that functions G_n and G are defined on a closed interval, that G is continuous, that each G_n is monotone decreasing, and that $G_n \to G$ pointwise as $n \to \infty$. Then $G_n \to G$ uniformly.*

We close Appendix 2 by a proof of Lemma 9.

Proof. Assume without loss of generality that the closed interval is [0, 1]. Pick $\epsilon > 0$. Since G is continuous on [0, 1], it is well known that G is uniformly continuous on [0, 1]. It is clear that G is monotone decreasing since each G_n is. Find an integer m so large that if x_1 and x_2 are points in [0, 1], then

$$| x_1 - x_2 | \leqslant 1/m \Rightarrow | G(x_1) - G(x_2)| < \epsilon. \tag{A2.8}$$

Find N so large that whenever $n > N$,

$$| G_n(x) - G(x)| < \epsilon \qquad \text{for} \quad x = 0, 1/m, 2/m,..., 1. \tag{A2.9}$$

We will show that if x in [0, 1] is arbitrary, and $n > N$, then $G(x) - 2\epsilon < G_n(x) < G(x) + 2\epsilon$, which proves the lemma. Find an integer k $(0 \leqslant k < m)$ such that $k/m \leqslant x \leqslant (k + 1)/m$. Write $a = k/m$ and $b = (k + 1)/m$. Then

$$\begin{aligned} G_n(x) &\geqslant G_n(b), && \text{since } G_n \text{ is monotone decreasing,} \\ &> G(b) - \epsilon, && \text{by (A2.9),} \\ &> G(x) - 2\epsilon, && \text{by (A2.8),} \end{aligned} \tag{A2.10}$$

and,

$$\begin{aligned} G_n(x) &\leqslant G_n(a), && \text{since } G_n \text{ is monotone decreasing,} \\ &< G(a) + \epsilon, && \text{by (A2.9),} \\ &< G(x) + 2\epsilon, && \text{by (A2.8).} \end{aligned} \tag{A2.11}$$

From (A2.10) and (A2.11) we obtain $G(x) - 2\epsilon < G_n(x) < G(x) + 2\epsilon$, as desired. ∎

APPENDIX 3

In this appendix, we will prove the following lemma, which is utilized in the proof of Theorem 1.

LEMMA 10. *Let B be a closed, bounded set. Then $(1 - (b/n)^{\tau_0 n} \to e^{-b\tau_0}$ as $n \to \infty$, uniformly over all b in B.*

Proof. Pick $\epsilon > 0$. Let $B_1 = \{e^{-b}: b \in B\}$. By uniform continuity of the function $x \to x^{\tau_0}$ on closed bounded sets, we can find $\delta > 0$ such that

$$| y - x | < \delta \text{ and } x \in B_1 \quad \Rightarrow \quad | y^{\tau_0} - x^{\tau_0} | < \epsilon. \tag{A3.1}$$

We will now show that

$$(1 - (b/n))^n \to e^{-b} \qquad \text{as } n \to \infty, \text{ uniformly over } b \text{ in } B. \tag{A3.2}$$

This is sufficient to prove the result of the lemma, because if n is sufficiently large, then

$$|(1 - (b/n))^n - e^{-b} | < \delta, \qquad \text{for all } b \text{ in } B \text{ by (A3.2)},$$

and hence

$$|(1 - (b/n))^{n\tau_0} - e^{-b\tau_0} | < \epsilon, \qquad \text{by (A3.1), where } y = (1 - (b/n))^n \text{ and } x = e^{-b}.$$

Now (A3.2) follows by Dini's theorem [10, p. 121] if we only show that there is an N such that for all $n > N$ and for all b in B, $(1 - (b/n))^n$ is a monotone increasing function of n. It is sufficient to show that $(1 - (b/x))^x$ has a positive derivative for $x > b$; take $N = \sup B$. The derivative is easily found to be

$$\begin{aligned}
&\left(1 - \frac{b}{x}\right)^{x-1} \left[\left(1 - \frac{b}{x}\right) \ln \left(1 - \frac{b}{x}\right) + \frac{b}{x}\right] \\
&\quad = \left(1 - \frac{b}{x}\right)^{x-1} \left[\left(1 - \frac{b}{x}\right)\left(-\frac{b}{x} - \frac{b^2}{2x^2} - \frac{b^3}{3x^3} - \frac{b^4}{4x^4} \cdots\right) + \frac{b}{x}\right] \\
&\quad \text{by using the Taylor series of } \ln \left(1 - \frac{b}{x}\right) \\
&\quad = \left(1 - \frac{b}{x}\right)^{x-1} \left[\left(1 - \frac{1}{2}\right) \frac{b^2}{x^2} + \left(\frac{1}{2} - \frac{1}{3}\right) \frac{b^3}{x^3} + \left(\frac{1}{3} - \frac{1}{4}\right) \frac{b^4}{x^4} + \cdots\right] \\
&\quad > 0. \quad \blacksquare
\end{aligned}$$

Acknowledgments

The author is grateful to Malcolm Easton for his helpful suggestions and discussions.

References

1. A. V. Aho, P. J. Denning, and J. D. Ullman, Principles of optimal page replacement, *J. Assoc. Comput. Mach.* 18 (1971), 80–93.
2. R. P. Boas, "A Primer of Real Functions," Carus Mathematical Monographs No. 13, Mathematical Association of America, Washington, D.C., 1960.
3. P. J. Denning and G. S. Graham, "Multiprogrammed Memory Management," IEEE Proceedings on Interactive Computer Systems, June 1975.
4. P. J. Denning and S. C. Schwartz, Properties of the working-set model, *Commun. ACM* 15 (1972), 191–198.

5. M. C. EASTON, Model for interactive data base reference string, *IBM J. Res. Develop.* **19** (1975), 550–556.
6. M. C. EASTON, Private communication.
7. R. FAGIN AND M. C. EASTON, The independence of miss ratio on page size, *J. Assoc. Comput. Mach.* **23** (1976), 128–146.
8. W. FELLER, "An Introduction to Probability Theory and its Applications," Vol. 1, 2nd ed., Wiley, New York, 1957.
9. E. GELENBE, A unified approach to the evaluation of a class of replacement algorithms, *IEEE Trans. Computers* **C-22** (1973), 611–618.
10. L. M. GRAVES, "The Theory of Functions of a Real Variable," 2nd ed., McGraw–Hill, New York, 1956.
11. W. P. HEISING, Note on random addressing techniques, *IBM Systems J.* **2** (1963), 112–116.
12. W. F. KING, III, "Analysis of Paging Algorithms," IFIP Conf. Proc. Ljubljana, Yugoslavia, August 1971.
13. D. E. KNUTH, "The Art of Computer Programming," Vol. 3, Addison–Wesley, Reading, Mass., 1973.
14. J. LENFANT, Doctoral dissertation, Rennes, France, 1974.
15. R. MATTSON, J. GECSEI, D. SLUTZ, AND I. TRAIGER, Evaluation techniques for storage hierarchies, *IBM Systems J.* **9** (1970), 78–117.
16. B. G. PRIEVE AND R. S. FABRY, "Evaluation of a Page Partition Replacement Algorithm," Bell Laboratories Technical Report, Naperville, Ill., October 1973.
17. B. G. PRIEVE AND R. S. FABRY, VMIN—an optimal variable-space page replacement algorithm, *Commun. ACM* **19** (1976), 295–297.
18. D. R. SLUTZ, "A Relation between Working Set and Optimal Algorithms for Segment Reference Strings," IBM Research Report RJ 1623, San Jose, Calif., July 1975.
19. J. H. WIMBROW, A large-scale interactive administrative system, *IBM Systems J.* **10** (1971), 260–282.
20. P. C. YUE AND C. K. WONG, On the optimality of the probability ranking scheme in storage applications, *J. Assoc. Comput. Mach.* **20** (1973), 624–633.
21. G. K. ZIPF, "Human Behavior and the Principle of Least Effort," Addison–Wesley, Cambridge, Mass., 1949.