W. F. Freiberger
U. Grenander
P. D. Sampson

Patterns in Program References

Abstract: This paper describes a study of some of the characteristics of program referencing patterns. Program behavior is investigated by constructing stochastic models for the page reference mechanism and evaluating the validity of the assumptions made through comparison with empirical results. The notion of a regime process is shown to play a useful role in describing the observed phenomena mathematically. The study falls within the realm of a rapidly growing field of computer science known as compumetrics, where quantitative and qualitative methods are being applied to the study and evaluation of computer performance.

Introduction

The execution of a program in a multiprogrammed system has to be interrupted frequently for reference to information stored in different levels of the storage hierarchy. To discuss the strategy of the basic decision algorithm appropriate for the system, it is necessary to know something about the manner in which these references to stored information are made.

When dealing with a large program, such as an assembler or a compiler, it is impossible, in practice, to predict the references deterministically, and it has been recognized for a long time that one has to resort to probabilistic models in this context.

The choice of the correct probabilistic model is far from obvious. It is the purpose of this paper to elucidate the problem by considering an analytic model and to relate the results of the analysis to actual measurements. Our aim consists in improving our understanding of the stochastic structure underlying the phenomena, but not yet to suggest methods for improvements in the decision algorithms of existing operating systems. The latter will be possible once the model has been firmly established and validated.

The model proposed in this paper should be regarded as only a first approximation to the best one. We could no doubt have obtained higher accuracy by choosing the stochastic process appearing in the next section to be more general, but this seemed to us not to be called for at the present preliminary stage of our investigations.

We start from some simple hypotheses about program references, in order to arrive at a model that is more intrinsic than pure curve-fitting would be:

1. Under multiprogramming, each program is given a "slice" of execution time that cannot be exceeded but may well not be used up. The size of the time slice can vary considerably among different systems, but will normally be enough for many thousands of operations. When we change from one program to another the probabilistic structure can also be expected to change, at least in an environment with a heterogeneous load. These breakpoints, and their distribution in time, will play an important role in any realistic model. Some other work [1-3] is in progress to develop statistical methods for studying this problem.
2. Between two breakpoints one can expect more homogeneous behavior, both because programs are executed sequentially, except for branching, and because the simplest kind of string information is also sequential. One might compare this with the notion of locality [4].
3. Branching makes for less homogeneous behavior and causes one of the most obvious aspects of program behavior, viz., looping. This will lead to an (approximately) periodic appearance, and with loops within loops there can be many periods present.
4. On the other hand, most programs have a characteristic behavior at their beginning (and usually also at their end). Initialization will be needed to set up tables, specify parameters, deal with macros and subroutines, etc. This also leads to heterogeneous behavior.

There are, of course, many other phenomena of program behavior that are known empirically although not yet quantified. For the moment, however, we shall limit ourselves to those mentioned above.

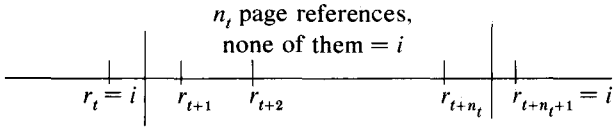
To be more specific, we consider an operating system that employs the concept of paging from and to virtual

memory. Pages are assumed to be of constant size and the entire storage hierarchy (main memory and auxiliary storage) in which paging takes place is taken to contain the collection $P = \{1, 2, 3, \cdots, M\}$ of pages, where M is the total number of pages. Let m be the maximum number of pages that can reside in main memory and denote the set of these pages by $\{1, 2, 3, \cdots, m\}$.

During execution the operating system makes references to memory and we shall denote by $p = \{\cdots, p_{-1}, p_0, p_1, \cdots\}$ the *reference string* of successive pages (not necessarily distinct) referred to. Under demand paging, a page that is referenced but is not resident in main memory will be brought in from auxiliary storage, usually in place of some other page which has to be pushed out, according to the particular paging algorithm implemented in the operating system. A common class of such algorithms is the so-called stack algorithm, of which the most popular is LRU (Least Recently Used); let us define stack algorithms as in [4].

The set $D = \{1, 2, \cdots, J\}$ denotes the set of pages of a given program, so that the members of the reference string $p_k \in J$. The program has been allocated a main memory space of m pages, where $1 \leq m \leq J$. We call a subset S of D such that S contains m or fewer pages a possible *memory state*. Let r be a reference string and A an allocation algorithm, and let $S(A, m, r)$ denote the memory state after A has processed r pages under demand paging in an initially empty main memory of size m. Then A is a stack algorithm if $S(A, m-1, r)$ is a subset of $S(A, m, r)$. That is, the contents of the $(m-1)$-page memory are always contained in the m-page memory, so that the memory states are "stacked up" on one another.

After considering this definition, it is clear why, for such stack algorithms, performance is crucially dependent on the behavior of the *distance string*, defined as follows. Suppose that at time t page $r_t = i$ has been referenced, and that the next reference to page i occurs at time $t + n_t + 1$ (see below); that is, between these two references to page i there have been n_t page references, but none to page i. Let d_t denote the number of *distinct* page references among these n_t references. Then the string $d = (\cdots, d_{-1}, d_0, d_1, \cdots)$ is called the *distance string* corresponding to the reference string r.



A page exception will occur each time that $d_t = m$.

Consider a substring $(r_a, r_{a+1}, r_{a+2}, \cdots, r_b)$ of the reference string and denote by $N_{a,b}$ the number of pages that had to be brought in from auxiliary storage during the interval $[a, b]$. The ratio $R_{a,b} = N_{a,b}/(b-a+1)$ is called the paging rate and is one of the criteria to be evaluated when judging the performance of a paged computing system. Notice that $R_{a,b}$ is an empirical quantity, not a parameter of any model.

It should be mentioned, in passing, that paging rates can be measured in at least two ways: either with respect to the flow of instructions (the unit being a time interval between two successive instruction executions) or with respect to *new* page references. Because the first method is more informative when studying overhead caused by page faults, we adopt it in this paper.

Model of program references

Let us measure time, as indicated above, in terms of the number of instructions executed. We shall not be concerned with happenings during individual instruction executions, but rather with nonoverlapping and contiguous time intervals of length T, where T is the *window* of the *working set*: the working set $W(t)$ is the binary M-vector with its ith component equal to one if the ith page has, zero if it has not, been referenced in the tth window $[(t-1)T, tT]$. In other words, $W(t)$ can be looked upon either as a binary vector or as the set whose indicator function is this distance.

The number of elements in $W(t)$ is

$$\#[W(t)] = \sum_{i=1}^{M} e_i(t) = \|W(t)\| \text{ (the Hamming norm)},$$

where $e_i(t)$ is the indicator function of page i.

Let us make the time parameter t continuous and treat $W(t)$ in terms of the "birth and death process" $e_i(t)$ (considering one i only):

$$e(t) = \left\{\begin{matrix}1\\ \text{or}\\ 0\end{matrix}\right\} \text{with} \left\{\begin{matrix}\text{mortality } \mu \text{ per time unit}\\ \text{and}\\ \text{birthrate } \lambda \text{ per time unit}\end{matrix}\right.$$

Here the time unit will be equal to the window size and the birthrate is equivalent to the reentry rate of pages.

We have the conditional probabilities

$$P_1(t) = P[e(t) = 1 | e(0) = 1], \text{ and}$$

$$P_0(t) = \mathrm{P}[e(t) = 1 | e(0) = 0].$$

Omitting subscripts since both P obey the same differential question but with different boundary conditions, we have (see [5], p.*459)

$$\begin{cases} P(t+h) = P(t)(1-\mu h) + [1-P(t)]\lambda h + o(h), \\ [P(t+h) - P(t)]/h = \lambda - P(t)(\mu+\lambda) + o(1), \\ (dP/dt) + (\lambda+\mu)P = \lambda, \\ P(t) = c\exp[-(\lambda+\mu)t] + \lambda/(\lambda+\mu). \end{cases}$$

We must have that $P_0(t) \to 0$ as $t \to 0$ and $P_1(t) \to 1$ as $t \to 0$; therefore

$$\begin{cases} P_0(t) = [1/(\lambda+\mu)](\lambda - \lambda \exp[-(\lambda+\mu)t]), \\ P_1(t) = [1/(\lambda+\mu)](\lambda + \mu \exp[-(\lambda+\mu)t]), \end{cases}$$

and we get the equilibrium probability $(t \to \infty)$

$$P[e(t) = 1] = \lambda/(\lambda+\mu) = p,$$

say, which holds of course for each page:

$$P[e_i(t) = 1] = \frac{\lambda_i}{\lambda_i + \mu_i} = p_i.$$

The equilibrium probability $(t \to \infty)$ for the size of the working set can be denoted by $P[w(t) = k] = w_k$, say; the whole probability distribution for the working set is denoted by $\{w\}$. Let $b(\alpha)$ denote the Bernoulli distribution with parameter α; then $\{w\}$ will be the convolution of M distributions with $\alpha = \lambda_i/(\lambda_i + \mu_i)$:

$$\{w\} = \prod_{i=1}^{M} {}^{*}\, b[\lambda_i/(\lambda_i + \mu_i)].$$

The generating function of $\{w\}$ is

$$\prod_{i=1}^{M} (1 - p_i + p_i z) = \exp\left\{\sum_{i=1}^{M} \ln(1 - p_i + p_i z)\right\}$$
$$= \exp\{M \int \ln(1 - p + pz)\ U(dp),$$

where $U(p)$ is the distribution function of the p_i.

The covariance function of each of the $e_i(t)$ is

$$\begin{aligned} \mathrm{Cov}\{e(t)\} &= \mathrm{E}[e(t)e(0)] - p^2 \\ &= [p/(\lambda+\mu)]\{\lambda + \mu \exp[-(\lambda+\mu)t]\} - p^2 \\ &= [\lambda\mu/(\lambda+\mu)^2] \exp[-(\lambda+\mu)t], \end{aligned}$$

so that the total spectral density becomes

$$\frac{1}{2\pi} \sum_{\nu} \frac{\lambda_\nu \mu_\nu}{(\lambda_\nu + \mu_\nu)^2[1 + (\lambda+\mu)^2\lambda^2]},$$

because

$$\exp(-|\alpha|t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(it\lambda)\,d\lambda}{|1 - i\alpha\lambda|^2}.$$

It should be noted that the resulting spectral density is monotonically decreasing for $\lambda \geq 0$.

The frequency function of the working set size $w(t)$ can take many forms. If the p_i are all small it will be approximately Poisson; if the p_i are all close to one it will be approximately inverse Poisson; and if all p_i are near the center of the interval (0, 1) it will be approximately normal. Convolutions among these three cases can occur, so that the resulting frequency function can take many forms.

We have, however, the following result, which limits the possible qualitative appearance of the distribution.

Assume that for each page 0, 1, 2, 3, $\cdots$, $N-1$ we have a probability π^k of being referenced in a window of given size. Assume also independence between pages. This obviously cannot be strictly true. We expect, however, stochastic dependence in time, which we take into account, to be more important than dependence in space, i.e., between pages. We do realize, however, that our assumption requires additional empirical support, and as will be seen in a later section the empirical results do not always support our assumptions. Then the number of pages called in a window will be a stochastic variable $NP = b_0 + b_1 + \cdots + b_{N-1}$, where the b_k are independent Bernoulli variables $b_k = b(\pi^k)$. For certain special cases we can obtain closed-form expressions or approximations for the distribution of NP: binomial, Poisson, inverse Poisson, normal. Because these do not seem to occur in practice, one is led to ask whether it is possible to say something general and qualitative about the distribution. The question is partially answered by the following theorem.

Theorem Let NP have the frequency function $p_k = P(NP = k)$; $k = 0, 1, \cdots, N-1$. Then the ratios p_k/p_{k-1} are nonincreasing. In other words, the frequency function is logarithmically convex.

Proof Introduce the ratios $r_k^{\,t} = p_k^{\,t}/p_{k-1}^{\,t}$; $k = 1, 2, \cdots, t$ and assume that for a fixed value of t the sequence $r_k^{\,t}$ is nonincreasing in k. Since

$$p_k^{\,t+1} = p_k^{\,t}(1 - \pi^t) + p_{k-1}^{\,t}\pi^t,$$

we have

$$\begin{aligned} r_k^{\,t+1} &= \frac{p_k^{\,t}(1-\pi^t) + p_{k-1}^{\,t}\pi^t}{p_{k-1}^{\,t}(1-\pi^t) + p_{k-2}^{\,t}\pi^t} \\ &= \frac{r_k^{\,t}(1-\pi^t) + \pi^t}{1 - \pi^t + \pi^t/r_{k-1}^{\,t}}. \end{aligned}$$

Because $r_k^{\,t}$ is nonincreasing we have

$$r_k^{\,t}(1-\pi^t) + \pi^t \leq r_{k-1}^{\,t}(1-\pi^t) + \pi^t,$$
$$1 - \pi^t + \pi^t/r_{k-1}^{\,t} \geq 1 - \pi^t + \pi^t/r_{k-2}^{\,t}.$$

Dividing these inequalities by each other, we obtain

$$r_k^{\,t+1} \leq r_{k-1}^{\,t+1}.$$

On the other hand, we have, for $t = 2$,

$$p_0^{\,2} = (1-\pi^0)(1-\pi^1),$$
$$p_1^{\,2} = (1-\pi^0)\pi^1 + \pi^0(1-\pi^1),$$
and $$p_2^{\,2} = \pi^0\pi^1,$$

so that

$$r_2^{\,2} = p_2^{\,2}/p_1^{\,2} = \frac{\pi^0\pi^1}{(1-\pi^0)\pi^1 + \pi^0(1-\pi^1)}, \text{ and}$$

$$r_1^{\,2} = p_1^{\,2}/p_0^{\,2} = \frac{(1-\pi^0)\pi^1 + \pi^0(1-\pi^1)}{(1-\pi^0)(1-\pi^1)}.$$

This implies that $r_2^2 \leq r_1^2$ since this is the same as

$$\pi^0\pi^1(1-\pi^0)(1-\pi^1) \leq [(1-\pi^0)\pi^1 + \pi^0(1-\pi^1)]^2,$$

which inequality holds. This proves the assertion.

As a consequence we know that the frequency function of the sum $b_0 + b_1 + \cdots + b_{N-1}$ must be of one of the following forms: 1) nonincreasing, 2) nondecreasing up to a point and then nonincreasing, or 3) nondecreasing. This severely limits the theoretical shape of the distribution under the conditions we have imposed: U-shape and bimodal shape are impossible, just to mention two cases.

One consequence of this theorem is that the distribution of the working set size $w(t)$ will be unimodal. (Bimodal distributions, however, were found in the empirical results as reported in a later section, and therefore the assumptions of the theorem cannot be wholly satisfied in practice.) Another significant consequence is the relation

$$\mathrm{Var}(w) = \sum p_\nu(1-p_\nu) \leq \sum p_\nu = \mathrm{E}[w];$$

in other words, the working set size distribution $w(t)$ has either subnormal or normal, but never supernormal, variation.

The working set $W(t)$ itself can be looked upon as a stochastic process taking *sets* as its values. In addition to the function $w = \|W(t)\|$ we shall consider the following Boolean functions, which are physically meaningful:

$$I(t) = W(t) \wedge [\sim W(t-1)],$$

representing the pages referenced in window t but not in $t-1$. We have

$$\mathrm{E}\|I(t)\| = \sum (1-p_\nu)\lambda_\nu = \sum (\mu_\nu\lambda_\nu/\lambda_\nu + \mu_\nu).$$

Similarly, $O(t) = (\sim W(t)) \wedge W(t-1)$ means that the pages were referenced in window $t-1$ but not in window t. The expected size of $O(t)$ is the same as that of $I(t)$ since

$$\mathrm{E}\|O(t)\| = \sum p_\nu\mu_\nu - \sum (\lambda_\nu\mu_\nu/\lambda_\nu + \mu_\nu).$$

This is obvious because the model is stationary and we must have balance. Also, let $L(t) = \{W(t) = W(t-1)\}$, where the second equality sign should be read as a Boolean function. The locality is expressed by $L(t)$; its expected size

$$\mathrm{E}\|L(t)\| = \sum [p_\nu(1-\mu_\nu) + (1-p_\nu)(1-\lambda_\nu)]$$

$$= \sum \frac{\lambda_\nu + \mu_\nu - 2\lambda_\nu\mu_\nu}{(\lambda_\nu + \mu_\nu)}$$

also expresses the locality. Just as above, one can also get expressions for the spectral densities of the sizes of $O(t)$, $I(t)$, and $L(t)$.

This simple model can only be expected to apply in a *homogeneous regime*, as in the present section.

An informal definition of a *regime process* starts from a random mechanism producing the time points where regimes start and end and also describes the transitions between regimes at these points. Conditioned by the break points the process formulates a stationary stochastic process between two such time points where the conditional probability distribution may depend upon the first random mechanism. For a precise definition of the notion of regime process and its mathematical consequences see [6].

A regime that exhibits looping, as described in the next section, would require a modified model with $\lambda_\nu^{(t)}$, $\mu_\nu^{(t)}$ periodic functions of t. This, however, is a matter of the time scale chosen. If the period of the outer loop is of the order of the window size T, then the statement in the last sentence can be expected to hold. On the other hand, if the period is much smaller, the looping will not be very pronounced and the λ_ν, μ_ν can be left as constants. Finally, if the period is much larger than T it may be sufficient to treat the regime as time-homogeneous unless other program events interfere.

The periodic case will look rather different from the purely homogeneous regime. The distribution $\{w\}$ will be made up of *contributions belonging to the different phases of the regime*. Note that this composition will have the effect of a mixture, not of a convolution. Hence there need be no tendency toward the normal distribution because the central limit theorem is not operative. We expect two or more peaks in the frequency function, and the spectral density need no longer be decreasing, but can have several peaks.

The effect of initialization (see the last section) is difficult to anticipate in general, and the same is true of time intervals with a gradual transition between regimes. In a later section we show through measurements how some actual programs can behave.

Distribution of distance

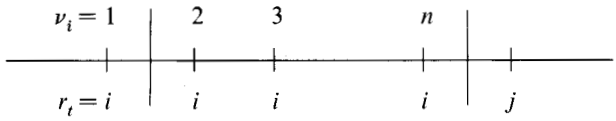
To study the behavior of the distance string we use the following model, which is a discrete-time version of the one used in the previous section.

Let $p = (p_1, p_2, \cdots, p_M)$ be a discrete probability distribution of the total number M of pages; i.e., $0 \leq p_i \leq 1$ and $\Sigma_{i=1}^M p_i = 1$. Also, let $q = (q_1, q_2, \cdots, q_M)$ be a vector with components which satisfy $0 \leq q_i \leq 1$. Our model develops as follows. At time t, a page i is selected from the collection of pages $P = \{1, 2, \cdots, i, \cdots, M\}$ according to the probability measure $p = \{p_1, p_2, \cdots, p_i, \cdots, p_M\}$. We now stipulate that subsequent references are made to this page i a number $\nu_i - 1$ times, where ν_i has the geometric probability distribution

$$P(\nu_i = n) = (1-q_i)q_i^{n-1}, \quad n = 1, 2, \cdots.$$

This means that q_i is the probability of staying within

page i and $(1-q_i)$ that of leaving it. It must be admitted that this assumption is far from convincing and is made for mathematical convenience. (A more resonable approximation to real data would be offered by the negative binomial distribution, which could be analyzed by appealing to Erlang's method of formulating the problems in terms of Markov chains.) In the diagram, pages between the long vertical bars are selected according to the probability measure $(1-q_i)q_i^{n-1}$; the last page before the first long vertical bar and the first page after the second long vertical bar are selected according to the probability measure p.



Note that

$$\sum_{n=1}^{\infty}(1-q_i)q_i^{n-1} = (1-q_i)\sum_{n=1}^{\infty}q_i^{n-1} = (1-q_i)\sum_{k=0}^{\infty}q_i^{k}$$

$$= (1-q_i)\cdot\frac{1}{1-q_i} = 1$$

as required. The expected value of the random variable ν_i is therefore

$$\mathrm{E}[\nu_i] = \sum_{n=1}^{\infty} n(1-q_i)q_i^{n-1} = \frac{1-q_i}{q_i}\sum_{n=1}^{\infty} n\,q_i^{n}$$

$$= \frac{1-q_i}{q_i}\cdot\frac{q_i}{(1-q_i)^2} = \frac{1}{1-q_i}$$

$$= m_i, \text{ say.}$$

After ν_i occurrences of page i we sample again from $P = \{1, 2, \cdots, i, \cdots, M\}$ according to the probability measure $p = \{p_1, p_2, \cdots, p_i, \cdots, p_m\}$ and proceed as before. Note that we may, of course, obtain the same page i again; i.e., in the above diagram we could have $j = i$.

Locality is large when q_i is close to unity, and hence when m_i is large.

We shall speak of the p_i as the *page frequencies* and of the m_i as the *page durations*. If the p-distribution is well concentrated, the work load is *skewed* toward a subset of pages; if the vector $m = \{m_i\}$ is large in magnitude, the work load possesses a high degree of locality.

This simple model may be criticized on at least two grounds: First, it is stationary; it is, however, known that this property is meaningful only in the context of time scale and can therefore not be ruled out a priori. Second—and this objection is more serious, the model postulates simple random sampling from $P = \{1, 2, \cdots, M\}$ as each new page is selected. Empirical studies are needed, and are reported on in a later section, to verify this assumption.

To relate the vectors $p = (p_1, p_2, \cdots, p_M)$ and $q = (q_1, q_2, \cdots, q_M)$ to data, the following considerations are relevant.

Let π_i be the equilibrium (or absolute) probability of referring to page i at some given moment: $P(r_t = i) = \pi_i$. We can write

$$\pi_i = \sum_j \pi_j\, p_{ji},$$

where p_{ji} is the transition probability of selecting page i at time t after page j has occurred at time $t-1$ in our (Markovian) model.

The transition from any state at time $t-1$ to state i at time t can, in our Markovian model, happen in three ways:

1. $r_{t-1} = r_t = i$ and we stay within an i substring (with probability q_i);
2. $r_{t-1} = r_t = i$ and we leave an i substring (with probability $1-q_i$) but select state i again (with probability p_i);
3. $r_{t-1} = j$, $r_t = i$ and we leave a j substring (with probability $1-q_j$), selecting a new state $i \neq j$ (with probability p_j).

Let the equilibrium (total) probabilities be π_j and the transition probabilities be p_{ji}. We have then, separating the above three cases,

$$\pi_j = \sum_j \pi_i\, p_{ji}$$

$$= \pi_i q_i + \pi_i(1-q_i)p_i + \sum_{j\neq i}\pi_j(1-q_j)p_i. \qquad (1)$$

Introducing the notation $a_i = \pi_i(1-q_i)$ and $a = \Sigma_i\, a_i$, we get from Eq. (1),

$$(1-p_i)a_i = p_i\sum_{j\neq i}\mathrm{a}_j = p_i(a-a_i),$$

so that $a_i = p_i a$, and

$$\pi_i = [p_i/1-\mathrm{q}_i)]a = p_i\, m_i\, a.$$

The constant a must be determined from the condition $\Sigma_i\,\pi_i = 1$, so that $a = (\Sigma\, p_k m_k)^{-1}$, and finally

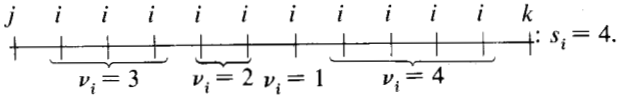
$$\pi_i = p_i m_i \Big/ \sum p_k m_k. \qquad (2)$$

Let M_i denote the expected length of a substring consisting entirely of i's; we have, for a single substring,

$P(\nu_i = n) = (1-q_i)q_i^{n-1}$, and

$\mathrm{E}[\nu_i] = m_i = 1/(1-q_i)$.

Let s_i = number of connected i substrings, e.g.,



Thus, $P(s_i = s) = p_i^{s-1}(1-p_i)$ and

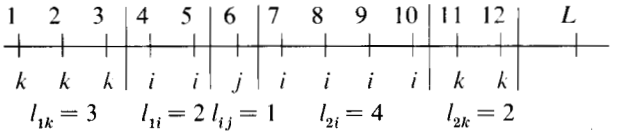
$$E[s_i] = \sum_{s=1}^{\infty} sP(s_i = s) = \sum_{s=1}^{\infty} s p_i^{s-1}(1-p_i),$$

$$= (1-p_i) \sum_{s=1}^{\infty} s p_i^{s-1} = 1/(1-p_i).$$

Hence,

$$M_i = m_i E[s_i] = m_i/(1-p_i). \tag{3}$$

Denote by L the length of a long reference string, by f_i the frequencies of page references to page i, and by $l_{i1}, l_{i2}, \cdots, l_{in_i}$ the lengths of the n_i connected i-substrings: i.e., l_{ij} will be the length of the jth connected i-string; e.g.,



Then, with convergence in probability, $f_i/L \to \pi_i$, and

$$(l_{i1} + l_{i2} + \cdots + l_{in_i}) n_i^{-1} \to M_i. \tag{4}$$

But Eqs. (2) and (3) give

$$\pi_i/M_i = p_i(1-p_i)/\Sigma p_k m_k. \tag{5}$$

The unknown parameter vectors p and q will not always be identifiable. We have, however, $\pi_i = p_i(1-p_i)M_i a$ where $a = (\Sigma p_k m_k)^{-1}$ so that, with some constant $C = 1/a$,

$$p_i(1-p_i)M_i = \pi_i C. \tag{6}$$

If we assume that all $p_i \leq 1/2$, which will certainly be true in all practical cases, then the p_i are uniquely determined from Eq. (6), choosing the smaller of the two roots. We select that (unique) value of C giving unity for the sum of the two roots p_i; this sum is clearly a monotonically increasing function of C. Then the relation (3) determines the other parameters m_i.

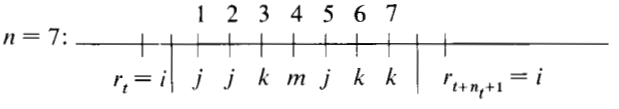
If all p_i are quite small, we can use the approximations

$m_i \cong M_i$, and

$P_i \cong (\pi_i/M_i)(\Sigma \pi_i/M_i)^{-1}$,

estimating M_i and π_i from Eq. (4). In other cases, it will be necessary to use an iterative procedure, starting with Eq. (3).

For the model defined in this section let us next determine the probability distribution of n_t, that is, of the length of the reference string between two equal page references $r_t = r_{t+n_t+1} = i$. Consider, first, the conditional probability $P(n_t = n | r_t = i)$; e.g.,



The conditional generating function of this probability is

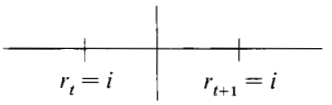
$$g_i(z|i) = E[z^{n_t} | i] = \sum_{n=0}^{\infty} e^{zn} P(n_t = n|i)$$

$$= q_i \cdot 1 + (1-q_i) E[z^{n_t} | i, \quad i\text{-string has ended}].$$

The first term corresponds to the situation $n = 0$ with $p_{t+1} = i$ being an element of an i-substring:

$z^n P(n_t = n|i) = z^0 P(n_t = 0|i) = 1 \cdot q_i$;

i.e.,



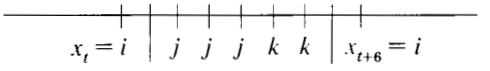
where the r_{t+1} is selected by the q_i mechanism.

We proceed to condition the expectation further and obtain

$$g_i(z|i) = q_i + (1-q_i) \sum_{l=0}^{\infty} p_i(1-p_i)^l$$

$$\times E[z^{n_t} | i, \ i\text{-string has ended}, \ l \text{ non-}i\text{-strings until } i];$$

e.g.,



Let $\gamma_i(z)$ denote the conditional generating function of the number ν of elements in one substring, given it is not an i-substring. Then

$$g_i(z|i) = q_i + (1-q_i) \sum_{l=0}^{\infty} p_i(1-p_i)^l \gamma_i^l(z)$$

$$= q_i + (1-q_i)p_i/[1-(1-p_i)\gamma_i(z)],$$

where

$$\gamma_i(z) = \sum_{j \neq i} \frac{p_j}{(1-p_i)} \sum_{\nu=1}^{\infty} (1-q_j) q_j^{\nu-1} z^{\nu}$$

$$= \frac{1}{(1-p_i)} \sum_{j \neq i} p_j(1-q_j) z \sum_{\nu=1}^{\infty} q_j^{\nu-1} z^{\nu-1}$$

$$= \frac{z}{(1-p_i)} \sum_{j \neq i} p_j(1-q_j) \frac{1}{(1-zq_j)}.$$

Clearly, the absolute generating function of the stochastic variable n_t now becomes

$$g(z) = \sum_{i=1}^{M} \pi_i g_i(z|i),$$

and we can evaluate the moments of n_t:

$$E[n_t] = g'(1) = \sum_{i=1}^{M} \pi_i g_i'(1|i).$$

Now,

$$g_i'(1|i) = -(1-q_i)\, p_i \frac{-(1-p_i)\gamma_i'(1)}{[1-(1-p_i)\gamma_i(1)]^2}.$$

Note that

$$\gamma_i(1) = [1/(1-p_i)] \sum_{j\neq i} p_j = 1,$$

as expected, and

$$\gamma_i'(1) = \sum_{j\neq i} [p_j(1-q_j)/(1-p_i)][1/(1-q_j) + q_j/(1-q_j)^2] = \sum_{j\neq i} p_j/(1-q_j)(1-p_i).$$

Hence,

$$g_i'(1/i) = [(1-q_i)/p_i] \sum_{j\neq i} p_j/(1-q_j)$$

and

$$\mathrm{E}[n_t] = \sum_{i=1}^{M} \frac{p_i m_i}{\sum_k p_k m_k} \frac{(1-q_i)}{p_i} \sum_{j\neq i} \frac{p_j}{1-q_j}$$

$$= \frac{1}{\sum_{k=1}^{M} p_k/(1-q_k)} \sum_{i=1}^{M} \sum_{j\neq i} \frac{p_j}{(1-q_j)};$$

because $m_i = \dfrac{1}{1-q_i}$,

$$\mathrm{E}[n_t] = \Big[\sum_{i,j=1, i\neq j}^{M} p_j/(1-q_j)\Big] \sum_{i=1}^{M} p_i/(1-q_i).$$

Summing over the entire $M \times M$ square matrix and subtracting the diagonal elements gives

$$\mathrm{E}[n_t] = \sum_i \sum_j p_j/(1-q_j) = \sum_i p_i/(1-q_i) \times \Big[\sum_j p_j/(1-q_j)\Big]^{-1}$$

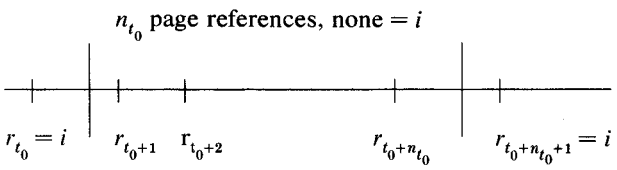
$$= M - \left(\frac{\sum_i p_i/(1-q_i)}{\sum_i p_i/(1-q_i)}\right) = M-1.$$

(It has been brought to our attention that this result can also be obtained more directly by an appeal to Markov Chain Theory.)

The distribution of the distance d_t is more complicated. To find its expected value we write

$$d_t = b_1 + b_2 + b_3 + \cdots + b_M,$$

where $b_l = 0$ or 1 according to whether page 1 has or has not been referenced.



As we see in diagram above, we have for the conditional expectation of b_l, given reference $r_t = i$ for $t = t_0$ and that $n_{t_0} = n$ selections are made from $P = \{1, 2, \cdots, M\}$ before i is referenced again (at $t = t_0 + n + 1$),

$$\mathrm{E}[b_l | r_{t_0} = i, n \text{ selections from } P \text{ before } i \text{ referenced again}] = 1 - (1-p_l')^n,$$

where we have introduced the conditional probability

$$P(r_t = j;\ j \neq i) = p_j' = p_j/(1-p_i);$$

$$j = 1, 2, \cdots, i-1, i+1, \cdots, M.$$

We thus get

$$\mathrm{E}[b_l] = \sum_{\substack{i=1 \\ (i\neq l)}}^{M} \mathrm{E}[b_l | r_{t_0} = i, n \text{ selections} \neq i) \times P(r_{t_0} = i) \times P(\text{exactly } n \text{ selections after } t_0 \text{ which are} \neq i)].$$

However,

$$P(r_{t_0} = i) = \pi_i,$$

and

$$P(\text{exactly } n \text{ selections after } t_0 \text{ which are} \neq i) = \sum_{n=0}^{\infty} p_i(1-p_i)^n(1-q_i).$$

Therefore

$$\mathrm{E}[b_1] = \sum_{i\neq 1} \sum_{n=0}^{\infty} \pi_i\, p_i (1-p_i)^n\, [1-(1-p_l')^n](1-q_i),$$

$$= \sum \pi_i\, p_i \left\{\sum_{n=0}^{\infty} (1-p_i)^n - \sum_{n=0}^{\infty} [(1-p_i)(1-p_l')]^n\right\}(1-q_i).$$

But

$$(1-p_i)(1-p_l') = (1-p_i)[1-p_l/(1-p_i)] = (1-p_i-p_l).$$

Therefore

$$\mathrm{E}[b_l] = \sum_{i\neq l} \pi_i\, p_i \left[\frac{1}{1-(1-p_i)} - \frac{1}{1-(1-p_i-p_l)}\right](1-q_i)$$

$$= \sum_{i \neq l} \pi_i \left[1 - \frac{p_i}{(p_i + p_l)}\right](1 - q_i) = \sum_{i \neq l} \frac{\pi_i p_l (1 - q_i)}{(p_i + p_l)}.$$

Hence, with $\pi_i = p_i m_i / \Sigma p_k m_k$,

$$\mathrm{E}[d_t] = \sum \mathrm{E}[b_l] = \sum_{\substack{i,j=1 \\ i \neq j}}^{M} \left[\frac{p_i p_j}{(p_i + p_j)}\right] \left(\sum_{k=1}^{M} p_k m_k\right)^{-1}.$$

Yue and Wong have obtained similar results [7].

If all pages had the same duration (m_i = constant), this last equation would reduce to the simple expression

$$\mathrm{E}[d_t] = \sum_{\substack{i,j=1 \\ i \neq j}}^{M} p_i p_j / (p_i + p_j).$$

In this case (m_i = constant) we have

$$\max \mathrm{E}[d_t] = (M - 1)/2$$

with the maximum realized for uniform page frequencies $p_i = 1/M$. To prove this, it is enough to use $xy/(x + y) \leq (x + y)/4$ with equality only if $x = y$. Then

$$\sum_{i \neq j} \frac{p_i p_j}{p_i + p_j} \leq \tfrac{1}{4} \sum_{i,j} (p_i + p_j) - \tfrac{1}{2} \sum_i p_i = \frac{M - 1}{2}, \quad \text{Q.E.D.}$$

If is, of course, natural that a uniform page frequency distribution will give maximum expected distance.

Leaving, now, the case of equal page durations (m_i = constant), we note that since $b_l^2 = b_l$ for each l, we have

$$\mathrm{E}[b_l^2] = \mathrm{E}[b_l].$$

To get the cross products we use, for l, l', i all different,

$$\mathrm{E}[b_l b_{l'} | r_t = i;\ k \text{ selections}] = 1 - (1 - p_l')^k - (1 - p_{l'})^i + (1 - p' - p_{l'}')^k,$$

so that

$$\begin{aligned}\mathrm{E}[b_l b_{l'}] &= \sum_{i \neq l, l'} \pi_i \left\{\sum_{i=0}^{\infty} [1 - (1 - p_l')^k - (1 - p_{l'})^k \right. \\ &\quad \left. + (1 - p_l' - p_{l'}')^k](1 - p_i)^k p_i\right\}, \\ &= \sum_{i \neq l, l'} \pi_i [1 - p_i/(p_i + p_l) - p_i/(p_i + p_l') \\ &\quad + p_i/(p_i + p_l + p_{l'}),\end{aligned}$$

which can be used to get an expression for the variance of the distance.

If M is large and all the p_i are small, the stochastic dependence between the b_l would be weak and it may therefore be tempting to apply the central limit theorem. If this argument were valid, the distance d_t should be asymptotically normally distributed. It will, however, be demonstrated below that this conculsion is, in general, false.

To do this, let us assume, for the moment, a uniform distribution $p_i = 1/M$, $q_i = 0$. The probability that n pages have been referenced, conditioned by 1) $r_t = i$ and 2) that k selections have been made, is denoted by

$$p(n, k) = P(n \text{ pages referenced, conditioned as stated});$$

$$n = 0, 1, \cdots, M - 1;$$

$$k = 0, 1, 2, \cdots$$

We can characterize $p(n, k)$ through a *partial difference equation*

$$p(n, k) = p(n, k - 1)\ n/(M - 1) + p(n - 1, k - 1) \times (M - n)/(M - 1),$$

for $k = 1, 2, \cdots$ and $n = 1, 2, \cdots, M - 1$. Introduce the transform (not a generating function!)

$$t_n(z) = \sum_{k=0}^{\infty} p(n, k)\ z^k;$$

substituting for $p(n, k)$ from above, and summing, yields

$$t_n(z) = p(n, 0) + [n/(M - 1)]\ z\ t_n(z) + [(M - n)/(M - 1)]\ z\ t_{n-1}(z).$$

But

$$p(n, 0) = \delta_{n0},\ t_0(z) = 1,$$

so that after some manipulations we get the solution

$$t_n(z) = \left(\prod_{\nu=1}^{n} \frac{M - \nu}{M - 1 - \nu z}\right) z^n.$$

Introducing the probabilities $P(n) = P(d_t = n)$ and using the fact that k has a geometric distribution with ratio $(M - 1)/M$, we get

$$P(n) = \sum_{k=0}^{\infty} \frac{1}{M} \left(\frac{M - 1}{M}\right)^{k-1} p(n, k) = \frac{1}{M}\, t_n\left(\frac{M - 1}{M}\right).$$

Substituting for $t_n\left(\frac{M - 1}{M}\right)$ from above, we get

$$P(n) = 1/M;\ n = 0, 1, \cdots, M - 1.$$

This obviously represents a far-from-Gaussian distribution; indeed, it is a rectangular distribution over the set of integers $(0, 1, 2, \cdots, n - 1)$! This result has a surprisingly simple form and should, perhaps, admit of an intuitive explanation and, also, a simpler proof than that given above. A plausibility argument based on symmetry would give such an explanation but this, of course, would not constitute a proof.

A more straightforward but less elegant way of handling this problem would have been to approach it through the classical occupancy problem, see [8, p 38].

For the case with general p (but still with all $q_i = 0$) the approach via a partial difference equation will not work. If we were to enlarge the state space, this approach would be feasible in principle, but not in practice. We

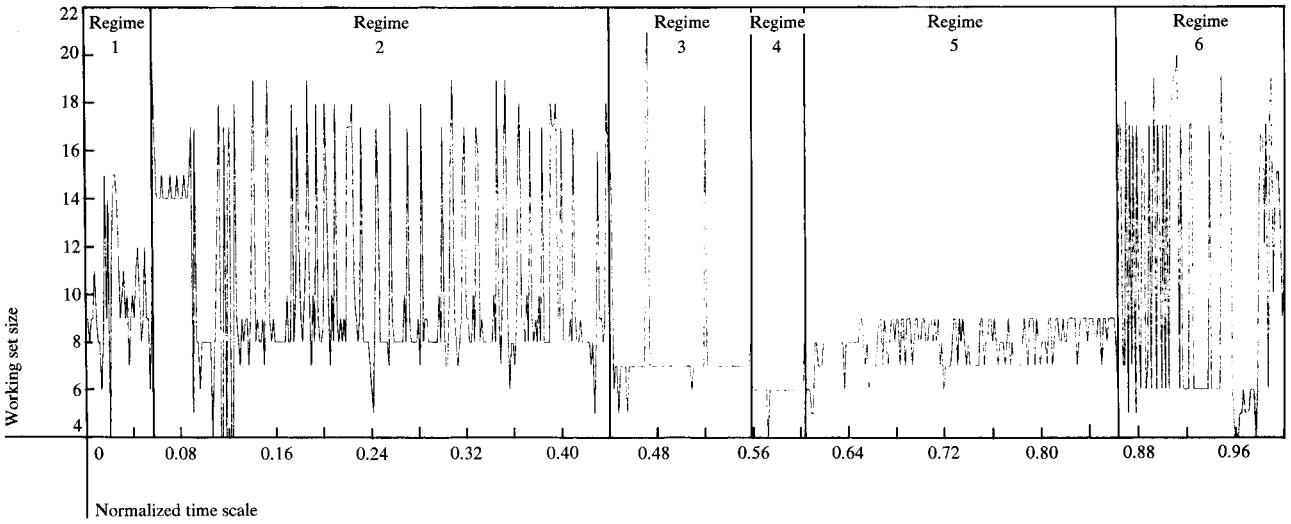


Figure 1 Graphical display of the six different regimes of the FORTRAN program MADD.

can, of course, write down an expression for $P(n)$ in terms of sums involving one, two, three, $\cdots p$'s at a time. The expression will have alternating signs and involve an enormous number of multiplications and additions.

Empirical results

In this section, we report some of the results of an experimental study of page reference patterns (for details, see [9]). This study clearly establishes the validity of the regime structure assumed in the model of the previous sections, although it suggests that some other assumptions—such as strict stochastic independence between pages—may not be completely valid.

We first describe the method we developed to partition page reference traces into homogeneous regimes. The reference traces were obtained from an interpretive simulator program [10] monitoring the performance of the execution of any program under the CP-67/CMS operating system. This monitoring program gathers information on the pages referenced at each instruction and records it in terms of the number of program references to each page per window of size T. The window size chosen was $T = 1000$ because we were not motivated by practical considerations of scheduler design but, rather, by a desire to understand the detailed stochastic structure underlying the page referencing mechanism.

The information was displayed on an interactive cathode ray tube device, and Fig. 1 shows an example of this display; it was also transferred to tape for processing. In view of our interest in the regime structure of large programs, we monitored the execution of a FORTRAN compiler designed to operate in about 128000 bytes of main storage and a PL/1 compiler designed to operate in about 64000 bytes. To determine the existence of stationary regimes, and to characterize these, we first produced listings of the references in each window of the 64 pages: a sample of such a listing for a FORTRAN compilation is shown in Fig. 2; the working set size for each window is given in the last column. This figure also illustrates our definition of stationary regimes as sets of windows in each of which the same set of pages (with few exceptions) is referenced for some minimum time interval. This definition is related to the concept of "locality of reference" [4]. In Fig. 2, the different regimes are partitioned by solid lines according to these heuristic guidelines. (The dashed lines indicate partitions arrived at through our algorithm described in the following section.) In this particular FORTRAN program (MADD) there are six distinct regimes. The same six regimes were found in all eight FORTRAN compilations examined and they clearly correspond to the six different phases of FORTRAN compilation [11]: Invoke, parse, allocate, unify, generate, and exit.

Figure 1 shows a plot against time of working set size for the same FORTRAN program, and the six regimes of stationary page referencing behavior are clearly displayed and divided by vertical lines. The histograms (sample frequency functions) for the regimes are displayed in Fig. 3 and the sample means, variances and first-order autocorrelations are given in Table 1. It is worth noting that similar histograms (some exhibiting bimodal behavior) are reported in [12]. The histograms suggest that the frequency functions for each regime have a characteristic shape peculiar to the corresponding phase of FORTRAN compilation. They also demonstrate that stochastic dependence between some pages must exist, contradicting one of the assumptions in the analytic model of the previous sections: It is shown there that if the pages are referenced independently, the frequency function of working set size must be logarithmic-

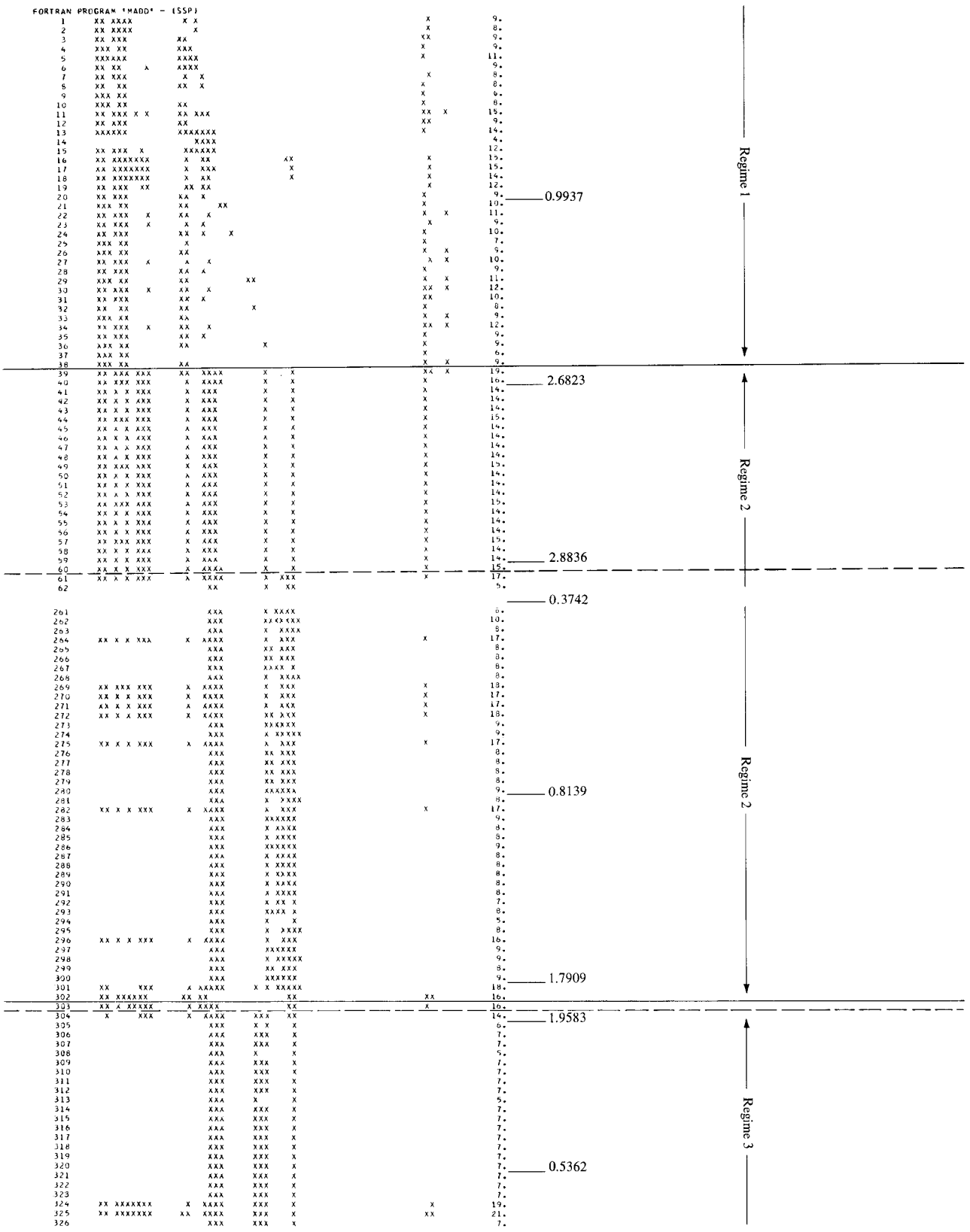


Figure 2 Illustration of the stationary regimes (listing form) of the FORTRAN program MADD.

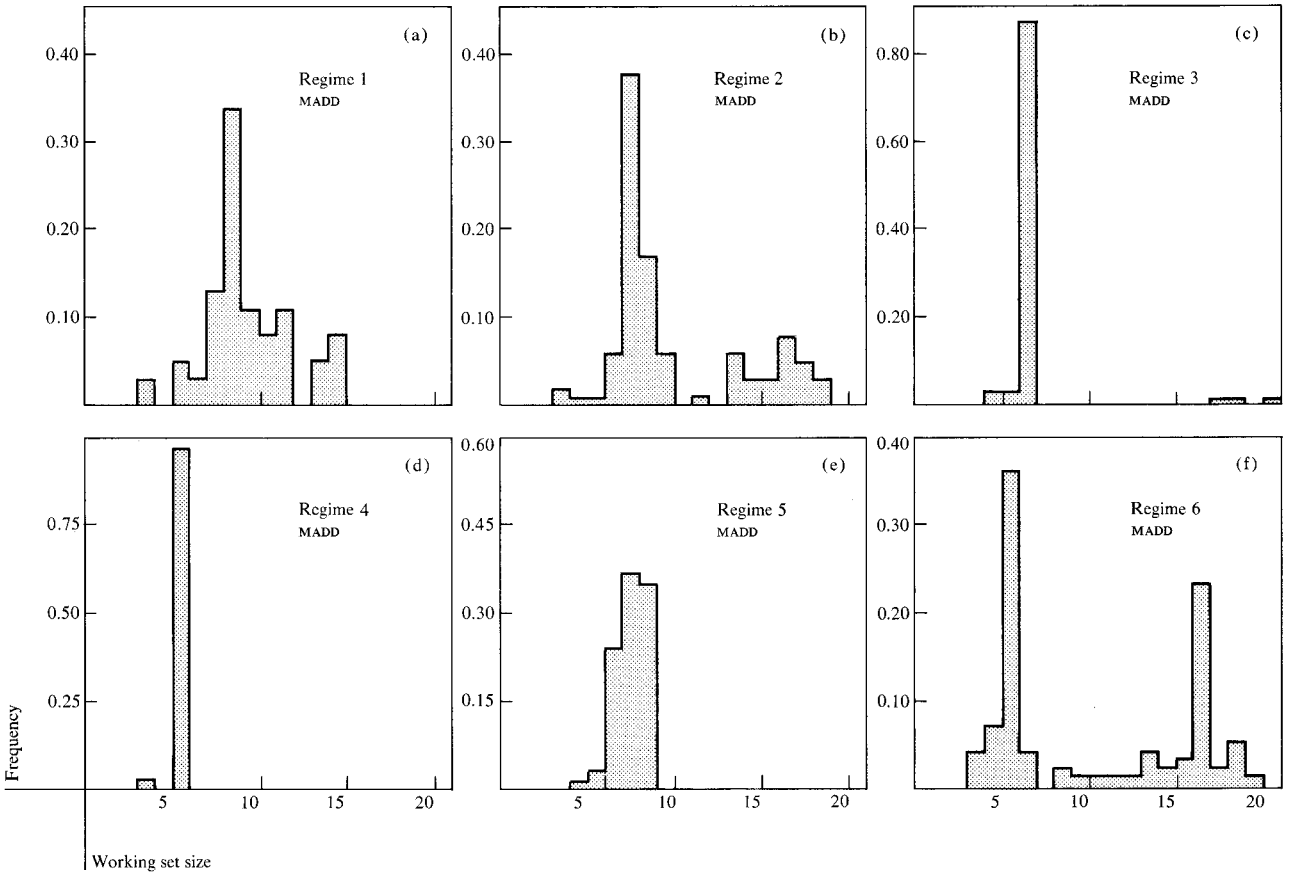


Figure 3 Histograms (sample frequency functions) of the FORTRAN program MADD.

ally convex and hence unimodal; regimes 2, 3 and 6 have, however, bimodal histograms. Table 1 shows means, variances and autocorrelations for the FORTRAN compilation of another program, PAGETR, in addition to MADD. (Other FORTRAN compilations gave similar results.) Corresponding regimes in the compilations show similar relations between means and variances, but not consistently enough for us to base a characterization of regimes on these parameters. The autocorrelation coefficients of order one indicate that (with the possible exception of the short regime 4) white noise cannot be the cause of the variation of working set size within regimes, because such a coefficient must, in the case of a time-series generated by white noise, be of the order $1/(n-1)^{1/2}$, where n is the number of observations [13]. In our case, thus, n is the number of windows in a given regime and almost all the sample autocorrelation coefficients are more than two standard deviations away from zero. Our results on PL/I compilations yielded similar though more complex results, because the number of observed regimes was higher and the working set sizes were larger.

Algorithm for partitioning regimes in page reference traces.

An algorithm was developed to partition page reference traces into homogeneous regimes in a manner reflecting the heuristic guidelines described above. We defined a stationary regime by a set of pages (or a locality of reference) common to almost all windows in the regime for some minimal length of time, that is, each of the 64 pages (including those not in the locality) is referenced with some characteristic frequency in each regime. Those pages that fall in the locality of reference defining a regime will of course have the highest frequencies. They will also carry the greatest weight in determining where regimes should be partitioned.

Because of the inter-page dependencies, noted in the previous section, the existence of these characteristic frequencies is reflected in the graphs of the sample frequency functions of the working set sizes in Fig. 3. That is, because of inter-page dependencies, each occurrence of a working set size (e.g., of size twelve) in a particular regime will usually be the result of references to the same

Table 1 FORTRAN compilation results.

Regime	*Program*	*Windows*	*Mean*	*Variance*	*First-order autocorrelation*
1	MADD	1-38	9.8421	6.3528	0.1671
	PAGETR	1-38	9.9211	4.5612	0.2591
2	MADD	39-300	10.4084	15.1620	0.2312
	PAGETR	39-594	9.5396	13.1379	0.1784
3	MADD	305-383	7.3924	5.8824	0.3342
	PAGETR	598-690	8.3763	12.2372	0.3783
4	MADD	385-414	5.9333	0.1333	−0.0344
	PAGETR	692-734	5.9535	0.1883	0.1060
5	MADD	416-592	8.0113	0.8294	0.3676
	PAGETR	736-1022	8.4913	1.2646	0.2066
6	MADD	594-688	10.6529	30.0376	0.2100
	PAGETR	1024-1138	11.6957	29.4240	0.0962

set of (twelve) pages, the majority of which (e.g., eight) will constitute the locality of references found in almost every window of the regime. Another working set size, corresponding to a different window in the same regime (e.g., fifteen), will usually reflect the same (eight) pages in the basic locality of reference and (seven) other pages different from the additional (four) pages in the windows with working set sizes of, e.g., twelve. With this situation in mind, we see that the observed characteristic shapes of the working set size frequency functions indicate characteristic frequencies of reference for each of the 64 pages.

Our algorithm is based on the assumption that a stationary regime of page referencing behavior can be characterized by a vector of 64 page-reference frequencies. It was remarked in the previous section that characteristic working set size frequencies and, consequently, characteristic frequency vectors correspond to compilations of different FORTRAN programs. This fact, which is significant, is, however, not essential to our algorithm which depends only on a marked change in frequency vectors between stationary regimes. Because of this, the algorithm is program-independent and should be capable of delineating stationary page reference patterns for any program displaying regime-type referencing behavior.

If a vector of 64 page reference frequencies can characterize a regime, then sample vectors of page reference frequencies within a regime should differ little over the span of the regime. Let $f^{(i)} = (f_1^{(i)}, \cdots, f_{64}^{(i)})$ be the sample frequency vector computed over the consecutive windows numbered $i, i+1, \cdots, i+N-1$. Let $f^{(i+N)} = (f_1^{(i+N)}, \cdots, f_{64}^{(i+N)})$ be the sample frequency vector computed over the windows numbered $i+N, i+N+1, \cdots, i+2N-1$. Then, if the $2N$ windows, $i, i+1, \cdots, i+2N-1$, all fall into the same regime, the mean square distance,

$$D = \|f^{(i)} - f^{(i+N)}\| = \left[\sum_{j=1}^{64} (f_j^{(i)} - f_j^{(i+N)})^2\right]^{1/2},$$

should be small. On the other hand, by assuming different localities of reference and different characteristic page reference frequencies in different regimes, if the windows $i, i+1, \cdots, i+N-1$ are contained wholly or mostly in one regime, and the windows $i+N, i+N+1, \cdots, i+2N-1$ are contained wholly or mostly in the next regime, then D should be large. In fact, D should be significantly larger only if the reference frequencies of a few pages change radically between regimes.

In accordance with the above remarks, our procedure is to compute the sample page reference frequency vectors over successive spans of N windows, searching for points in time where the mean square distance, D, between the frequency vectors is large. Suppose now that some sufficiently large value of D is found between frequency vectors computed over the windows $i-N, \cdots, i-1$, and $i, \cdots, i+N-1$. Then an attempt is made to maximize D as this should locate more precisely a partition point between stationary regimes. The frequency vectors are adjusted $2N$ times so that a value of D can be computed between N-vectors centered at each of the $2N$ windows $i-N, \cdots, i-1, i, i+1, \cdots, i+N-1$. If the maximum value of D is sufficiently large, then the particular window at which the maximum occurred is recognized as a partition point. In other words, it is assumed by this algorithm that if the distance between reference frequency vectors computed over consecutive N-window spans is maximized at window j, then the windows $j-N, \cdots, j-1$, and $j, \cdots, j+N-1$, will lie in entirely different

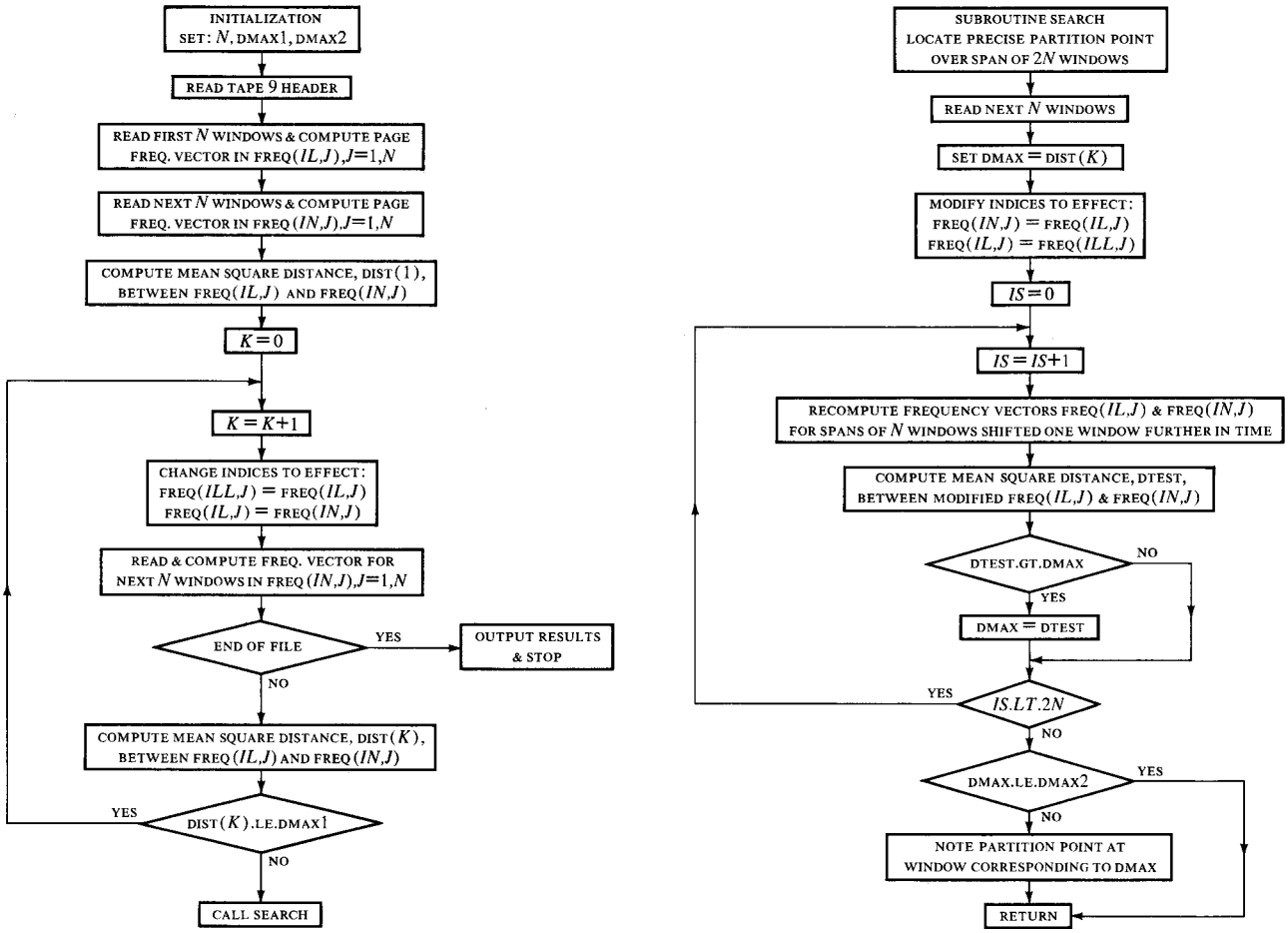


Figure 4 Flowchart of the algorithm used in the delineation of regimes.

regimes with different characteristic page reference frequency vectors. This is, however, an idealization as there is generally a transition period of at least a few windows between regimes. It is thus not possible to locate an exact partition point.

The algorithm just described is outlined in greater detail in the flow chart in Fig. 4. It was programmed in FORTRAN and, as indicated in the flow chart, it consists of a main program, REGIME, and a subroutine, SEARCH, which is called to locate precisely a partition point through a maximization of D.

Clearly the choice of the value of N, the number of windows over which sample frequency vectors are calculated, is vital to the success of the algorithm. The value must be large enough that a good estimate of the actual characteristic frequency vector is obtained. It must also be smaller than the length (in number of windows) of the shortest stationary regime that will be considered. Also important is the choice of a value of D "sufficiently" large. There are actually two such values. The first, DMAX1, is the value which must be exceeded to warrant a window-by-window search for a precise regime partition point. The second value, DMAX2, must be exceeded by the maximum value of D in the window-by-window search over $2N$ windows (in subroutine SEARCH) in order for a window to be recognized as a regime partition point.

The choice of the parameters N, DMAX1 and DMAX2 will depend on the nature of the particular class of programs being studied. The algorithm was consistently successful in partitioning the page reference traces for the FORTRAN compiler according to our design with the values $N = 20$, DMAX1 $= 0.99$ and DMAX2 $= 1.5$. This choice of N means that we will not recognize as stationary regimes any spans of fewer than 20 windows or 20000 instructions. Certainly, shorter spans of time are of little practical significance and a smaller value of N would make it difficult to estimate accurately the true characteristic frequency vector.

To the right of the computer printed output in Fig. 2 we have written the value of the distance D between successive frequency vectors in intervals of 20 windows.

The regime partition points for the two FORTRAN programs, as determined by this algorithm, have been marked by dashed lines in cases where these points differ from the heuristically selected partition points marked by solid lines. The maxima of the distance function corresponding to these partitions have been written above or below the dashed lines. Note that the minimum value of D at any window recognized as a partition point is 1.6771. In view of the transition periods between regimes, it is clear that this algorithm partitions regimes as accurately as could be expected.

We point out here that for the PL/I compilations, the best results were achieved with values of DMAX1 = 1.5 and DMAX2 = 2.5. These higher thresholds were necessitated by the generally higher working set sizes and values of D.

We emphasize that our algorithm is an approach to the problem of regime identification tailored to the problem at hand. Regime stochastic processes occur in many different applications, and different identification algorithms will be appropriate in each case. For more general discussions, see [1] and [6], and for different applications to compumetrics, [2] and [3].

Summary

In the theoretical part of this paper, the consequences of certain basic assumptions about the random nature of the page referencing mechanism in virtual memory operating systems is examined. Among the stochastic dependencies present in such a system are those between different time points and those between different pages. In one of our models, the latter is neglected in comparison with the former. Subsequent analytic studies in conjunction with experimental findings are shown to lead to a weakening of this hypothesis in favor of the notion of a regime stochastic process as an appropriate model for the description of empirical observations on page reference strings. An attempt is made, also, to correlate the observed regimes with the known structure of the executing program and to collect information on the properties of the internal regime processes, i.e., those processes conditioned by the one governing the regime transitions.

We look upon this work as a pilot study in the spirit of statistical inference applied to stochastic process models in compumetrics.

Acknowledgments

This work was partially supported by the Office of Naval Research, Information Systems Program, Contract N00014-67-A-0191-0020 and by a joint study contract with the IBM Cambridge Scientific Center. The authors are indebted to the reviewers for their valuable remarks.

References

1. Ben-Tung Ang, "Heuristic-Adaptive Search for Regimes in a String of Data," *Report of the Division of Applied Mathematics*, Brown University, July 1974.
2. G. Henderson and J. Rodriguez-Rosell, "The Optimal Choice of Window Sizes for Working Set Dispatching," *Report of the Division of Applied Mathematics*, Brown University, July 1974.
3. Wen-Te K. Lin, "A Statistical Study of an Operating System," *Report of the Division of Applied Mathematics*, Brown University, July 1974.
4. Peter J. Denning, "Virtual Memory," *Computing Surveys* **2,** 3 (1970).
5. William Feller, *An Introduction to Probability Theory and Its Application*, Volume II, John Wiley & Sons, Inc., New York, 1966.
6. R. K. Adenstedt, "Weather Regimes in Stochastic Meteorological Models," *Quarterly of Applied Mathematics* **28,** 343 (1970).
7. P. C. Yue and C. K. Wong, "On the Optimality of the Probability Ranking Scheme in Storage Applications," *J. ACM* **20,** 624 (1973).
8. William Feller, *An Introduction to Probability Theory and Its Applications*, Volume I, third edition, John Wiley & Sons, Inc., New York, 1968.
9. Paul D. Sampson, "Regime Behavior in Page Referencing Patterns of Computer Programs," *Report of the Division of Applied Mathematics*, Brown University, July 1974.
10. W. Millbrandt, "An Interpreter for Program References," *Report of the Division of Applied Mathematics*, Brown University, July 1974.
11. *IBM System/360 Operating System, FORTRAN IV(G) Compiler, Program Logic Manual*, Form GY22-6638-1, IBM Data Processing Division, 1133 Westchester Avenue, White Plains, New York 10604, and available at IBM branch offices.
12. Peter Bryant, "Predicting Working Set Sizes," *IBM J. Res. Develop.* **19,** 221 (1975); this issue.
13. M. S. Bartlett, *Stochastic Processes*, Cambridge University Press, London, 1955.

Received October 25, 1974; revised January 8, 1975

W. F. Freiberger and U. Grenander are located at the Division of Applied Mathematics, Brown University, Providence, Rhode Island. P. D. Sampson is now at the University of Michigan, Ann Arbor, Michigan 02912.