On the Bayesian approach to optimal performance of page storage hierarchies

By A. BENCZÚR, A. KRÁMLI, J. PERGEL

Introduction

In connection with the work of operating systems or interactive data base systems or large program systems of any other destination on computers with hierarchical memory arises the optimal page replacement problem. In two level storage hierarchy a reference to a page not in first level storage is called page fault. Optimal replacement algorithms minimize under different conditions the average number of page faults.

A great majority of papers devoted to this problem (see e.g. Aho et al. [1], Franaszek and Wagner [5], Easton [6]) assumes that the stochastic behaviour of the reference string is known. Therefore the algorithms proposed by them are only asymptotically optimal, when the probability distributions have to be estimated in the course of the execution of the program.

In this paper — using the Bayesian method, which first has been applied to this problem by Arató [7] — we prove in two extreme cases of loss function the optimality of the so called "least frequently used" strategy on every finite time interval for reference strings with unknown probability distribution.

Our considerations remember to the solution of the so called "two-armed bandit problem" (see Feldman [4]); we investigate the nature of the basic equation of dynamic programming (Bellman equation).

In § 1 we give the short description of the model and the formulation of the problem.

§ 1.

The program consists of n pages $1, 2, \ldots, n$, and m pages can be stored in the high speed memory and $n-m$ pages (often the whole program) are stored on a slow access memory device. The reference string $\{\eta_1, \ldots, \eta_t, \ldots\}$ from probabilistic point of view forms a sequence of independent identically distributed random variables, the common probability distribution

$$P_{i,w} = P_w(\eta_t = i)$$

of the random variables η_t depends on a parameter w, value of which is unknown. The dependence on w is given as follows: the range of parameter w is the set W of all permutations of natural numbers $1, \ldots, n$; $w(i)$ denotes the one to one mapping of set $\{1, \ldots, n\}$ realized by w. There is given a fixed decreasing sequence $p_1 > \ldots > p_n$ of probabilities $(p_1 + \ldots + p_n = 1)$ and $P_{i,w} = p_{w(i)}$.

Following the Bayesian approach to the decision theory we assume that w itself is a random variable — as we have no preliminary information about the distribution $P(\eta_t = i)$ the a priori distribution of parameter w is the uniform one.

Let us denote by $D_{t,N}$ the set of all possible sequential decision procedures $\{d_t, \ldots, d_{N-1}\}$ on a finite time interval $[t, N]$. A decision $d_{t'}$, which depends only on the initial decision d_0 and the observed reference string $\{\eta_1, \ldots, \eta_{t'}\}$ $(t' \in [t, N])$ means the subset of pages being absent of the central memory after the observation of string $\{\eta_1, \ldots, \eta_{t'}\}$.

The decision d_t consists of $n-m$ elements. By Arató's model (case A) the memory can be rearranged without extra cost before each reference η_t, but a page-fault $(\eta_t \in d_{t-1})$ increases the cost by 1 unity; i.e. the loss function has the following form

$$X_t^{d_{t-1}} = \begin{cases} 1 & \text{if } \eta_t \in d_{t-1}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this paper there is investigated another extreme case (case B) too: each change of a page increases the cost by 1 unity and η_t always must be stored in the central memory; i.e., the loss function has the following form

$$X_t^{d_t, d_{t-1}} = |d_t \setminus d_{t-1}|, \tag{2}$$

where $|.|$ denotes the number of elements of a finite set. (Notice that if $\eta_t \in d_{t-1}$, then $X_t^{d_t, d_{t-1}} \geqq 1$.)

§ 2. (Case A)

Our aim is to find the set of sequential decision procedures $\{d_0 \ldots d_{N-1}\}$ which minimize the risk function

$$E\left(\sum_{t=1}^{N} X_t^{d_{t-1}}\right)$$

(E is the expectation taken on the basis of the a priori distribution of w.) In the sequel y_t denotes the fixed value of η_t.

Let

$$v(y_1, \ldots, y_t, N-t) = \min_{\{d_t, \ldots, d_{N-1}\} \in D_{t,N}} E_{y_1, \ldots, y_t}\left(\sum_{\tau=t+1}^{N} X_\tau^{d_{\tau-1}}\right), \tag{3}$$

where $E_{y_1, \ldots, y_t}$ denotes the conditional expectation under a given string $\{y_1, \ldots, y_t\}$.

The class of functions $v(y_1, \ldots, y_t, N-t)$ satisfies the Bellman equation (see e.g. [2])

$$v(y_1, \ldots, y_{t-1}, N-t+1) = \min_{d_{t-1}} E_{y_1, \ldots, y_{t-1}}\left(X_t^{d_{t-1}} + v(y_1, \ldots, y_{t-1}, \eta_t, N-t)\right). \tag{4}$$

Solving it recursively we can find the set of optimal strategies. Notice that $v(y_1, \ldots, y_t, N-t)$ does not depend on d_{t-1}, therefore, it is sufficient to minimize for every t the conditional expectation

$$E_{y_1, \ldots, y_{t-1}}(X_t^{d_{t-1}}).$$

We shall prove that the optimal strategies are those for which d_0 is arbitrarily chosen and d_t consists of the $n-m$ least frequently occured pages in the string $\{y_1, \ldots, y_t\}$. Before this we prove a lemma and two corollaries of it.

Lemma 1. Let us suppose that the frequency f_i of the page i in the string $\{y_1, \ldots, y_t\}$ is less than the frequency f_j of the page j. Let w_1 and w_2 be two permutations of natural numbers $1, \ldots, n$, and $k_1 < k_2 \leqq n$ two natural numbers.

If

$$\begin{gathered} w_1(i) = k_1, \quad w_1(j) = k_2, \\ w_2(i) = k_2, \quad w_2(j) = k_1, \\ w_1(k) = w_2(k) \quad \text{for every} \quad k \neq i, j, \end{gathered} \tag{$*$}$$

then

$$P(w_1|y_1, \ldots, y_t) < P(w_2|y_1, \ldots, y_t).$$

Proof. On the basis of Bayes' theorem

$$P(w_1|y_1, \ldots, y_t) = \frac{\prod_{k=1}^{n} p_{w_1(k)}^{f_k}}{\sum_{w \in W} \prod_{k=1}^{n} p_{w(k)}^{f_k}}. \tag{5}$$

Similarly,

$$P(w_2|y_1, \ldots, y_t) = \frac{\prod_{k=1}^{n} p_{w_2(k)}^{f_k}}{\sum_{w \in W} \prod_{k=1}^{n} p_{w(k)}^{f_k}}. \tag{6}$$

The assertion of our Lemma can be obtained by comparison of (5) and (6) using the inequality $p_{k_2} < p_{k_1}$.

Corollary 1. If $\{y_{t+1}, \ldots, y_{t+\tau}\}$, $\{\bar{y}_{t+1}, \ldots, \bar{y}_{t+\tau}\}$ are two strings (sequences of pages) i, j are two pages with the following properties, for every $1 \leqq \tau' \leqq \tau$

(i) if $y_{t+\tau'} \neq i, j,$ then $\bar{y}_{t+\tau'} = y_{t+\tau'},$

(ii) if $y_{t+\tau'} = i,$ then $\bar{y}_{t+\tau'} = j,$

(iii) if $y_{t+\tau'} = j,$ then $\bar{y}_{t+\tau'} = i,$

(iv) the frequency f_i of the page i in the string $\{y_1, \ldots, y_t\}$ is less, than the frequency f_j of the page j,

(v) if the frequency of the page j in the string $\{y_{t+1}, \ldots, y_{t+\tau}\}$ is greater than the frequency of page i, then

$$P(\eta_{t+1} = y_{t+1}, \ldots, \eta_{t+\tau} = y_{t+\tau}|y_1, \ldots, y_t) > P(\eta_{t+1} = \bar{y}_{t+1}, \ldots, \eta_{t+\tau} = \bar{y}_{t+\tau}|y_1, \ldots, y_t). \tag{7}$$

Proof. The set W can be decomposed into the union of $\frac{n!}{2}$ disjoint pairs of permutations $\{w_1, w_2\}$ of $(*)$ property figuring in Lemma 1. Inequality (7) can be obtained by direct comparison applying Lemma 1 to every such pair.

Remark 1. Corollary 1 means for $\tau=1$ that the order of aposteriori probabilities of the pages after having observed the string $\{y_1, \ldots, y_t\}$ is the same as the order of their frequencies in this string.

Corollary 2. Let $\{y_{t+1}, \ldots, y_{t+\tau}\}$ and $\{\bar{y}_{t+1}, \ldots, \bar{y}_{t+\tau}\}$ be the same strings and i, j the same pages as in Corollary 1. Let A and B be two events of the algebra generated by random variables $\eta_{t+\tau+1}, \ldots, \eta_N$, which are invariant under the changing of i and j; let i' and j' be two pages different from i and j. If

$$A \setminus B = \{\eta_{t+\tau''} = i'\}, \quad B \setminus A = \{\eta_{t+\tau''} = j'\}$$

for a suitable τ'', then

$$\begin{aligned}
&|P(\eta_{t+1} = y_{t+1}, \ldots, \eta_{t+\tau} = y_{t+\tau}, A|y_1, \ldots, y_t)- \\
&-P(\eta_{t+1} = y_{t+1}, \ldots, \eta_{t+\tau} = y_{t+\tau}, B|y_1, \ldots, y_t)| > \\
&> |P(\eta_{t+1} = \bar{y}_{t+1}, \ldots, \eta_{t+\tau} = \bar{y}_{t+\tau}, A|y_1, \ldots, y_t)- \\
&-P(\eta_{t+1} = \bar{y}_{t+1}, \ldots, \eta_{t+\tau} = \bar{y}_{t+\tau}, B|y_1, \ldots, y_t)|.
\end{aligned} \tag{8}$$

The proof is analogous to the proof of Lemma 1 and Corollary 1. Inequality (8) can be obtained by comparison of conditional probabilities for every quadruple $\{w_1, w_2, w_3, w_4\}$ of permutations of the following property.

If $k_1, k_2, k_3, k_4 \leqq n$ are 4 different natural numbers, then

$$\begin{aligned}
&w_1(i) = k_1, \quad w_1(j) = k_2, \quad w_1(i') = k_3, \quad w_1(j') = k_4, \\
&w_2(i) = k_2, \quad w_2(j) = k_1, \quad w_2(i') = k_4, \quad w_2(j') = k_3, \\
&w_3(i) = k_1, \quad w_3(j) = k_2, \quad w_3(i') = k_3, \quad w_3(j') = k_4, \\
&w_4(i) = k_2, \quad w_4(j) = k_1, \quad w_4(i') = k_4, \quad w_4(j') = k_3,
\end{aligned}$$

and

$$w_1(k) = w_2(k) = w_3(k) = w_4(k) \quad \text{for every} \quad k \neq k_1, k_2, k_3, k_4.$$

Theorem 1. The set of sequential decision procedures $\{d_0, \ldots, d_{N-1}\}$ which minimize the expected loss $E\left(\sum_{t=1}^{N} X_t^{d_{t-1}}\right)$ in case A consists of the so called least frequently used (LFU) strategies, i.e. d_0 is arbitrary, and for every t, d_t consists of the first $n-m$ least frequently used pages in the string $\{y_1, \ldots, y_t\}$.

Proof. Theorem 1 is a straightforward consequence of Remark 1 and the uniformity of the a priori distribution of parameter w.

§ 3. (Case B)

In case B form we follow the method of comparing the expected cost of optimal continuations of different decisions d_t and d_t' after having observed the string $\{y_1, \ldots, y_t\}$. Therefore, we denote by $v(y_1, \ldots, y_t, d_t, N-t)$ the risk-function belonging to the observations $y_1, \ldots, y_t$ the state d_t of memory at time t and the optimal strategy on the time interval $[t+1, N]$, i.e.,

$$v(y_1, \ldots, y_t, d_t, N-t) = \min_{\{d_{t+1}, \ldots, d_{N-1}\} \in D_{t+1, N}} E_{y_1, \ldots, y_t}\left(\sum_{\tau=t+1}^{N} X_\tau^{d_\tau, d_{\tau-1}}\right).$$

Our aim is to determine the set of sequential decision procedures $\{d_0, \ldots, d_{N-1}\}$ for which

$$\min_{\{d_0, \ldots, d_{N-1}\} \in D_{0, N}} E\left(\sum_{t=1}^{N} X_t^{d_t, d_{t-1}}\right) = \min_{d_0} v(d_0, N)$$

is reached.

First we prove a lemma, which restricts the set of possible strategies to the so called demand paging algorithms.

Lemma 2. If $\eta_t \notin d_{t-1}$; d_t and d_t' are two different decisions of properties

(*i*) $d_t = d_{t-1}$,

(*ii*) $|d_t' \setminus d_t| = l > 0$,

then

$$v(y_1, \ldots, y_t, d_t, N-t) \leqq l + v(y_1, \ldots, y_t, d_t', N-t). \tag{9}$$

Proof. There are 4 possible cases

$$\eta_{t+1} \in d_t' \setminus d_t,$$

$$\eta_{t+1} \in d_t \setminus d_t',$$

$$\eta_{t+1} \in d_t \cap d_t',$$

$$\eta_{t+1} \in \{1, \ldots, n\} \setminus (d_t \cup d_t').$$

In every case it is easy to show that for an arbitrary decision d_{t+1}' the decision d_{t+1} can be chosen to be equal to d_{t+1}' paying at most l extra cost.

Remark 2. A similar assertion can be verified for $\eta_t \in d_{t-1}$. If d_t and d_t' are two different decisions with properties

(*iii*) $|d_t \setminus d_{t-1}| = 1$,

(*iv*) $|d_t' \setminus d_{t-1}| = l$,

(*v*) $|d_t' \setminus d_t| = l-1$,

then

$$v(y_1, \ldots, y_t, d_t, N-t) \leqq v(y_1, \ldots, y_t, d_t', N-t) + l - 1. \tag{10}$$

The Bellman equation for the risk-functions has the form

$$v(y_1, \ldots, y_{t-1}, d_{t-1}, N-t+1) =$$

$$= \min_{d_t} E_{y_1, \ldots, y_{t-1}} \left(X_t^{d_t, d_{t-1}} + v(y_1, \ldots, y_{t-1}, \eta_t, d_t, N-t) \right). \tag{11}$$

The relations (9) and (10), applying recursively equation (11), show that the optimal sequential decision procedures are among those which fulfil the conditions

$$d_t = d_{t-1} \quad \text{if} \quad \eta_t \in d_{t-1}$$

and

$$d_{t-1} \setminus d_t = \{\eta_t\} \quad \text{if} \quad \eta_t \in d_{t-1}. \tag{12}$$

The decision procedures of the above type are called "demand paging algorithms" (see e.g. Denning [3]).

We can deduce from the following theorem that the LFU strategies minimize the expected loss in case B, too.

Theorem 2. If d_t and d_t' are two different decisions for which

$$d_t \setminus d_t' = \{i\}, \quad d_t' \setminus d_t = \{j\}$$

and the frequency f_i of the page i in the string $\{y_1, \ldots, y_t\}$ is less than the frequency f_j of the page j, then

$$v(y_1, \ldots, y_t, d_t, N-t) < v(y_1, \ldots, y_t, d_t', N-t). \tag{13}$$

Proof. The proof can be carried out by induction on $\theta = N-t$. If $\theta = 1$, then the assertion of Theorem 2 is an obvious consequence of Corollary 1. Applying the induction hypothesis for $\theta = 1, \ldots, \theta = N-1-t$ we get that the optimal decisions in every case $\eta_t \in d_{t-1}$ are those for which $d_t \setminus d_{t-1}$ is one of the least frequently used (in the string $\{y_1, \ldots, y_t\}$) pages of the admissible set $\{1, \ldots, n\} \setminus d_{t-1}$.

To demonstrate the main idea of the proof of the induction step, first we briefly present it in the case $n=3$, $m=2$. Then $d_t = \{i\}$, $d_t' = \{j\}$. Let us denote by k the third page. If $f_k \geqq f_i$ then, using the induction hypothesis, it is easy to prove that for arbitrary outcome y_{t+1} of η_{t+1},

$$v(y_1, \ldots, y_{t+1}, d_{t+1}, N-t-1) \leqq v(y_1, \ldots, y_{t+1}, d_{t+1}', N-t-1), \tag{14}$$

where $d_{t+1}(d_{t+1}')$ is the optimal continuation of decision $d_t(d_t')$.
But

$$E_{y_1, \ldots, y_t}(X_{t+1}^{d_{t+1}, d_t}) \leqq E_{y_1, \ldots, y_t}(X_{t+1}^{d_{t+1}', d_t'})$$

by Corollary 1, thus using the equation (11) (Bellman equation) we get inequality (13).

If $f_k < f_i$, then in the case $\eta_{t+1}=j$ inequality (14) fails, therefore, we need to analyze the strings of the form

$$(a)\ \eta_{t+1}=i, \ldots, \eta_{t+\tau'-1}=i, \quad \eta_{t+\tau'}=k,$$

$$(b)\ \eta_{t+1}=i, \ldots, \eta_{t+\tau-1}=i, \quad \eta_{t+\tau}=j,$$

$$(a')\ \eta_{t+1}=j, \ldots, \eta_{t+\tau'-1}=j, \quad \eta_{t+\tau'}=k,$$

$$(b')\ \eta_{t+1}=j, \ldots, \eta_{t+\tau-1}=j, \quad \eta_{t+\tau}=i.$$

Using the same arguments as in case $f_k \geqq f_i$ it is easy to show that after having observed a string of type (*a*), (*a'*), (*b*) or (*b'*) the optimal continuation of decision d_t has a better or equal optimal continuation than those of decision d_t'. For cases (*a*) and (*a'*) the increments of conditional risks for the optimal continuations of the decisions d_t and d_t' can be compared using Corollary 1.

For the cases (*b*) and (*b'*) we have to prove the inequality

$$\begin{aligned} & P(\eta_{t+1}=i|y_1, \ldots, y_t)+P(\eta_{t+1}=j, \eta_{t+2}=i|y_1, \ldots, y_t)+\ldots \\ & \quad +P(\eta_{t+1}=j, \ldots, \eta_{N-1}=j, \eta_N=i|y_1, \ldots, y_t) \leqq \\ & \leqq P(\eta_{t+1}=j|y_1, \ldots, y_t)+P(\eta_{t+1}=i, \eta_{t+2}=j|y_1, \ldots, y_t)+\ldots \\ & \quad +P(\eta_{t+1}=i, \ldots, \eta_{N-1}=i, \eta_N=j|y_1, \ldots, y_t). \end{aligned} \tag{15}$$

Inequality (15) can be verified using Corollary 1 and the obvious relation

$$\begin{aligned} & P(\eta_{t+1}=i|y_1, \ldots, y_t)+P(\eta_{t+1}=j, \eta_{t+2}=i|y_1, \ldots, y_t)+\ldots+ \\ & +P(\eta_{t+1}=j, \eta_{t+2}\neq i, \ldots, \eta_{t+\tau-1}\neq i, \eta_{t+\tau}=i|y_1, \ldots, y_t)+\ldots= \\ & = P(\eta_{t+1}=i|y_1, \ldots, y_t)+P(\eta_{t+1}=j|y_1, \ldots, y_t)= \\ & = P(\eta_{t+1}=j|y_1, \ldots, y_t)+P(\eta_{t+1}=i, \eta_{t+2}=j|y_1, \ldots, y_t)+\ldots+ \\ & +P(\eta_{t+1}=i, \eta_{t+2}\neq j, \ldots, \eta_{t+\tau-1}\neq j, \eta_{t+\tau}=j|y_1, \ldots, y_t)+\ldots \end{aligned} \tag{16}$$

as (15) can be obtained from (16) by leaving pairs on term from the left hand side the other from the right hand side so that in each pair the left hand side term is greater.

Next we give the proof of the general case. Let us denote by I the subset of pages from the set $\{1, \ldots, n\}\setminus d_t'$ with less than f_i frequency in the string $\{y_1, \ldots, y_t\}$. If $|I|=0$, then the proof is analogous to that of case $f_k \geqq f_i$ for $n=3$.

When $\eta_{t+1}\neq i, j$, then $X_{t+1}^{d_{t+1}, d_t}=X_{t+1}^{d'_{t+1}, d'_t}$ and

$$v(y_1, \ldots, y_{t+1}, d_{t+1}, N-t-1) \leqq v(y_1, \ldots, y_{t+1}, d'_{t+1}, N-t-1) \tag{17}$$

by the induction hypothesis.

For $\eta_{t+1}=i$,

$$X_{t+1}^{d_{t+1}, d_t}=1, \quad X_{t+1}^{d'_{t+1}, d'_t}=0 \tag{18}$$

and

$$v(y_1, \ldots, y_t, i, d_{t+1}, N-t-1) < v(y_1, \ldots, y_t, i, d'_{t+1}, N-t-1). \tag{19}$$

Similarly, for $\eta_{t+1}=j$,

$$X_{t+1}^{d_{t+1}, d_t} = 0, \quad X_{t+1}^{d'_{t+1}, d'_t} = 1 \tag{20}$$

and by condition $|I|=0$

$$v(y_1, ..., y_t, j, d_{t+1}, N-t-1) = v(y_1, ..., y_t, j, d'_{t+1}, N-t-1). \tag{21}$$

By Corollary 1, from (18) and (20) follows the inequality

$$E_{y_1, ..., y_t}(X_{t+1}^{d_{t+1}, d_t}) < E_{y_1, ..., y_t}(X_{t+1}^{d'_{t+1}, d'_t}),$$

which together with (17), (19), (21) and the Bellman equation proves the assertion of Theorem 2 in case $|I|=0$.

Let us assume that $|I|\neq 0$, and introduce the mapping Φ on the set of strings $\{y_{t+1}, ..., y_{t+\tau}\}$ of length τ (τ is an arbitrary natural number, and $y_{t+\tau}\in\{1, ..., n\}$) as follows

$$\{\bar{y}_{t+1}, ..., \bar{y}_{t+\tau}\} = \Phi(\{y_{t+1}, ..., y_{t+\tau}\})$$

and for every $1\leqq\tau'\leqq\tau$,

$$\begin{aligned} \bar{y}_{t+\tau'} &= y_{t+\tau'} \quad \text{if} \quad y_{t+\tau'} \neq i, j, \\ \bar{y}_{t+\tau'} &= j \quad \text{if} \quad y_{t+\tau'} = j, \\ \bar{y}_{t+\tau'} &= i \quad \text{if} \quad y_{t+\tau'} = j. \end{aligned}$$

Let us investigate the behaviour of the optimal continuations of decisions d_t and d'_t on the strings of the form

$$\{y_{t+1} = j, y_{t+2} \neq i, ..., y_{t+\tau-1} \neq i, y_{t+\tau} = i\}.$$

There exists a $\tau''>1$, such that for $\tau'\leqq\tau''$ the following relations are valid

(i) $d_{t+\tau'}=d_{t+\tau'-1}$ or the unique element of $d_{t+\tau'}\setminus d_{t+\tau'-1}$ has less frequency in the string $\{y_1, ..., y_{t+\tau'}\}$ than the page i.

(ii) $d'_{t+\tau'}=d'_{t+\tau'-1}$ or the unique element of $d'_{t+\tau'}\setminus d'_{t+\tau'-1}$ has less frequency in the string $\{y_1, ..., y_{t+\tau'}\}$ than the page i.

(iii) $d_{t+\tau'}\setminus d'_{t+\tau'} = \{i\}$,

(iv) $d'_{t+\tau'}\setminus d_{t+\tau'} = \{k_1^*\}$

and there is at most one k_2^* in the set $\{1, ..., n\}\setminus d'_{t+\tau}$ which has less frequency in the string $\{y_1, ..., y_{t+\tau'}\}$ than the page k_1^*. The properties (*i*), (*ii*) and (*iii*) are obvious consequences of the induction hypothesis, the property (*iv*) can be proved by induction on τ'.

If the first moment τ' for which property (*ii*) fails, is less than τ, then

$$d'_{t+\tau'}\setminus d'_{t+\tau'-1} = \{i\} \quad \text{i.e.} \quad d'_{t+\tau'} = d'_{t+\tau}. \tag{22}$$

(Notice, that for such a τ' property (*i*) is still valid.)

Therefore,

$$v(y_1, \ldots, y_{t+\tau'}, d_{t+\tau'}, N-t-\tau') = v(y_1, \ldots, y_{t+\tau'}, d'_{t+\tau'}, N-t-\tau'). \tag{23}$$

If there is no such $\tau' < \tau$ for which property (*ii*) fails, and

$$d'_{t+\tau-1} \setminus d_{t+\tau-1} = \{k_1^*\}$$

has minimal frequency among the pages $\{1, \ldots, n\} \setminus d'_{t+\tau'}$, then it follows from the relation

$$y_{t+\tau} = i \quad \text{that} \quad d'_{t+\tau} = d_{t+\tau},$$

i.e.,

$$v(y_1, \ldots, y_{t+\tau}, d_{t+\tau}, N-t-\tau) = v(y_1, \ldots, y_{t+\tau}, d'_{t+\tau}, N-t-\tau).$$

If there exists a unique page k_2^* in the set $\{1, \ldots, n\} \setminus d'_{t+\tau}$ with less frequency than k_1^*, i.e.,

$$d'_{t+\tau} \setminus d_{t+\tau} = \{k_2^*\}, \quad d_{t+\tau} \setminus d'_{t+\tau} = \{k_1^*\}, \tag{24}$$

then we have to argue more carefully, which we shall do after having analyzed the behaviour of optimal continuations of decisions d_t and d'_t on the strings of type

$$\Phi(\{y_{t+1} = j, y_{t+2} \neq i, \ldots, y_{t+\tau-1} \neq i, y_{t+\tau} = i\}).$$

Let us denote by $\bar{d}_{t+\tau}$ and $\bar{d}'_{t+\tau}$ the optimal continuations of decisions d_t and d'_t on the string

$$\Phi(\{y_{t+1}, \ldots, y_{t+\tau}\}).$$

If for a $\tau' < \tau$, $d_{t+\tau'}$ and $d'_{t+\tau'}$ fulfil the conditions (*i*) and (*ii*) on the string $\{y_{t+1}, \ldots, y_{t+\tau'}\}$, then so do $\bar{d}_{t+\tau'}$ and $\bar{d}'_{t+\tau'}$. Moreover,

(*v*) $d'_{t+\tau'} = \bar{d}_{t+\tau'}$,

(*vi*) $\bar{d}'_{t+\tau'} \setminus d_{t+\tau'} = \{j\}$,

(*vii*) $d_{t+\tau'} \setminus \bar{d}'_{t+\tau'} = \{i\}$.

Obviously for $\eta_{t+1} = i$, $\bar{d}_{t+1}$ and $\bar{d}'_{t+1}$ satisfy the relation

$$v(y_1, \ldots, y_t, i, \bar{d}_{t+1}, N-t-1) < v(y_1, \ldots, y_t, i, \bar{d}'_{t+1}, N-t-1).$$

But this inequality is insufficient for the proof of Theorem 2, as we have to ballance the difference in expected loss between the continuation of decisions d_t and d'_t on the strings of type

$$\{y_{t+1} = j, \; y_{t+2} \neq i, \ldots, y_{t+\tau-1} \neq i, \; y_{t+\tau} = i\}.$$

By properties (*i*)—(*vii*), the symmetry of the mapping Φ and Corollary 1 the sum of expected loss of continuations of decisions d_t and $\bar{d}_t$ on the strings $\{y_{t+1}, \ldots, y_{t+\tau}\}$ and $\Phi(\{y_{t+1}, \ldots, y_{t+\tau}\})$ is less than those of decisions d'_t and $\bar{d}'_t$.

The comparison of probabilities of page faults at the moment τ caused by decisions $d_{t+\tau-1}$ and $\bar{d}'_{t+\tau-1}$ can be carried out analogously to the case $n=3$, using the identity (15). It remains to analyze the case when there exists a page k_2^* in the set $\{1, \ldots, n\} \setminus d'_{t+\tau}$ with less frequency than k_1^*. By the symmetry of the mapping Φ, for the string $\Phi(\{y_{t+1}, \ldots, y_{t+\tau}\})$,

$$\bar{d}_{t+\tau}\setminus\bar{d}'_{t+\tau}=\{k_2^*\}$$

and

$$\bar{d}'_{t+\tau}\setminus\bar{d}_{t+\tau}=\{k_1^*\} \tag{25}$$

(See properties (*iv*), (*v*) and relation (24).)

Let $\tau''>\tau$ the moment of first page-fault caused by decision $d_{t+\tau}$ or $d'_{t+\tau}$ (respectively by $\bar{d}'_{t+\tau}$ or $\bar{d}_{t+\tau}$) on the string $\{y_{t+1}, \ldots, y_{t+\tau}, y_{t+\tau+1}, \ldots, y_N\}$ (respectively on $\{\Phi(\{y_{t+1}, \ldots, y_{t+\tau}\}), y_{t+\tau+1}, \ldots, y_N\}$). Since k_2^* and k_1^* are the first two pages of the set

$$\{1, \ldots, n\}\setminus(d_{t+\tau}\cap d'_{t+\tau})=\{1, \ldots, n\}\setminus(\bar{d}_{t+\tau}\cap\bar{d}'_{t+\tau})$$

least frequently used in the string $\{y_1, \ldots, y_t, y_{t+1}, \ldots, y_{t+\tau}\}$ (respectively $\{y_1, \ldots, y_t, \Phi(\{y_{t+1}, \ldots, y_{t+\tau}\})\}$ thus we get

$$v(y_1, \ldots, y_{t+\tau''}, d_{t+\tau''}, N-t-\tau'')=v(y_1, \ldots, y_{t+\tau''}, d'_{t+\tau''}, N-t-\tau''), \tag{26}$$

$$\begin{aligned} v\big(y_1, \ldots, y_t, \Phi(\{y_{t+1}, \ldots, y_{t+\tau}\}), y_{t+\tau+1}, \ldots, y_{t+\tau''}, \bar{d}_{t+\tau''}, N-t-\tau''\big) &= \\ = v\big(y_1, \ldots, y_t, \Phi(\{y_{t+1}, \ldots, y_{t+\tau}\}), y_{t+\tau+1}, \ldots, y_{t+\tau''}, \bar{d}'_{t+\tau''}, N-t-\tau''\big). \end{aligned} \tag{27}$$

If the page fault was caused by an event of type

$$\eta_{t+\tau''}\in d_{t+\tau}\cap d'_{t+\tau},$$

then the expected loss of the strategy $\{\ldots d'_t, \ldots, d'_{t+\tau}, \ldots\}$ is greater than the loss of the other one. In the opposite case we can compare the common expected loss of the strategies

$$\{\ldots d'_t, \ldots, d'_{t+\tau}, \ldots\} \quad \text{and} \quad \{\ldots \bar{d}'_t, \ldots, \bar{d}'_{t+\tau}, \ldots\}$$

(respectively

$$\{\ldots d_t, \ldots, d_{t+\tau}, \ldots\} \quad \text{and} \quad \{\ldots \bar{d}_t, \ldots, \bar{d}_{t+\tau}, \ldots\})$$

using Corollary 2, and we get that the former is greater. This last remark together with relations (26) and (27) completes the proof of Theorem 2.

Remark 3. Also in case **B** the decision d_0 can be arbitrarily chosen by the symmetry of the a priori distribution of the parameter w.

Remark 4. Our all considerations remain valid for any a priori distribution in the space of all probability distributions invariant under the permutations of pages.

Abstract

Using the sequential Bayesian method the authors prove that in a two level storage hierarchy the "least frequently used" strategy is optimal for the page fault rate. It is assumed that the reference string forms a sequence of independent identically distributed random variables with unknown distribution. Two kind of loss function is discussed.

References

[1] AHO, A. V., P. J. DENNING, J. D. ULLMANN, Principles of optimal page replacement, *J. Assoc. Comput. Mach.*, v. 18, 1971, pp. 80—93.
[2] DE GROOT, M. H., *Optimal statistical decisions*, Mc Graw—Hill, N. Y., 1970.
[3] DENNING, P. J., Virtual memory, *Comput. Surveys*, v. 2, 1970, pp. 153—189.
[4] FELDMAN, D., Contributions to the "two-armed bandit" problem, *Ann. Math. Statist.*, v. 33, 1962, pp. 847—856.
[5] FRANASZEK, P. A., T. J. WAGNER, Some distribution-free aspects of paging algorithm performance, *J. Assoc. Comput. Mach.*, v. 21, 197 , pp. 31—39.
[6] EASTON, M. C., Model for interactive data base reference string, *IBM J. Res. Develop.*, v. 19, 1975, pp. 550—556.
[7] ARATÓ, M., A note on optimal performance of page storage, *Acta Cybernet.*, v.3, 1976, pp. 25—30.

(Received March 11, 1976)