Measuring, Modelling and Evaluating Computer Systems,
H. Beilner and E. Gelenbe, (eds.)

STATISTICAL SEQUENTIAL METHODS FOR UTILIZATION IN PERFORMANCE ANALYSIS

M. Arató
Research Institute for Applied
Computer Sciences
Budapest, Hungary

This paper is a survey of certain results concerning point processes and their model approximations in computer systems. We develop a unique method based on known statistical sequential models in order to compute automatic decision rules. Applications of this approach to time-sharing systems and performance analysis are presented. Exact results of Shiryayev and Davis are presented and their use in different approximations are shown.

The independent reference string models are also discussed to show that the separation principle gives computable results in dynamic file assignement and in page replacement too.

1. INTRODUCTION

The purpose of this paper is to present in the manner of statistical sequential decision procedures some of the fundamental results which are directly applicable to the performance analysis of multiprogramming systems. Here we do not use the well known Wald's sequential likelilood ratio conception for distinguishing between two hypothesis, which may be used in many applications of computer systems. Such an example the reader may find in this paper /see fig. 3./ where the overflow and underflow level for the number of allocated pages in a virtual memory system may be found from Wald's rule. For another example we may use the dispatching algorithm in the HASP execution task monitor in IBM OS/360 /see Shohat, Strauss [30] /.

The main goal of this paper is to show that the stopping rule procedures and recursive filtering of stochastic processes may be used in many problems of complex computer performance analyses. If we are interested in a set of service stations interconnected into a network only in "whole" then we can work out decision procedures for such measure of performance effectiveness as degree of multiprogramming, swapping rate, overhead rate, etc. /Some models and measurements the reader may find e.g. in Töke [31] and Asztalos [10]/. Taking into account e.g. the swapping rate we take a complex and stochastic function of the elements in the network which varies stochastically in time. In such a representation it is possible to control this process. The first question which arises is to detect the changes in the behavior of the network system. In this paper, we give the Poisson process description /see Davis [16] /, and the Wiener process case too /see Shiryayev [29] /. The Wiener process representation is used as an approximation applying an appropriate time scale transformation. Here we do not deal with the problem of recursive filtering the reader may find it in the papers of Segall [28], and Arató [7] . The disorder problem is one of the Bayesian sequential procedures which has an elegant solution. Until now this decision procedure was not used in internal work of computer systems. This is one way to analyse a computer system not as a static one. Generally, in sequential procedures we do not assume that the parameters of the system are known or they cannot vary in time. If the parameters of the system vary with time, or they are unknown, a dynamical treatment might give not only a substantial improvement in performance, but there exist also special cases when we have relatively simple solutions /see e.g. Arató, Benczúr, Krámli [8] , Benczúr, Krámli [12] /.

The problem of finding optimal paging algorithm with an independent reference string η_t $(t = 1,2,\dots)$ is well known /see Aho, Denning, Ullmann [1] , Bélády [13] or Gelenbe [19] /. A simple model of file assignement in computer networks with Markov request string was discussed by Segall [28] . These two models with independent reference strings were discussed on this conference in 1976. The method which we used was the Bayesian and the performance index was the expected number of page faults. We did not assume that the distributions are known /see Arató [5] , Benczúr, Krámli, Pergel [11] /. Later we proved the optimality of the LFU strategy without the Bayesian assumption /Arató, Benczúr, Krámli [8] /. This means that a result similar to Wald's theorem on optimality of the likelihood sequential procedure /where the expectation of sample time is minimal under both hypotheses/ is true. It is known that Wald's theorem on optimality has a more complicated proof than the Bayesian one.

In this paper we present in a unique method the separation principle for independent strings. The separation theorem for linear Gaussian models was proved by Wonham [33] /see also in Lipcer, Shiryayev [25]/. The separation principle means that the estimates of the unknown probability distribution are sufficient statistics for the optimization.

2. THE DISORDER PROBLEM

a/ In complex computer systems we have processors /e.g. central processor units, input-output channels, remote terminals, etc./ which we call service stations, and tasks, processes, messages /called transactions in information management system/, which we call customers. A set of these service stations are interconnected to each other by a topology, and this forms a network. Customers enter the network via sources and are directed to the service stations. A computer system may also be considered from another viewpoint: The processes /or tasks, messages/ could be the service stations which are utilizing a set of resources. The resources are created by the arrival process of customers and may be destroyed by the departure of customers. A resource may be obtained by a process after having been utilized by another process. The precise definition of these models with service stations and number of customers in the whole network $K = \sum_{i=1}^{n} k_i$ the reader may find in the paper of Gelenbe, Muntz [20] . Consider Figure 1. where an n service system is illustrated.

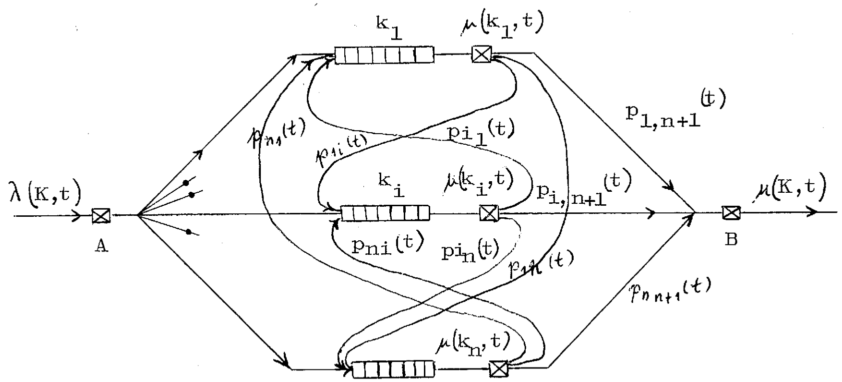

Figure 1.

In many cases in such a system we are interested only in the arrival process of customers at point A and the departure process at point B, or in a complex function of them /e.g. in a performance measure/. We may assume, as a first approximation that the analyzed stochastic processes are Poisson type and the rate $\lambda(K,t)$ /or $\mu(K,t)$ / depends on time and the total number of customers in the system. The parameters $\mu(k_i)$ and $p_{ij}(t)$ mean the departure rate at i-th service station and transition probability from i-th service station to j-th at time t, respectively.

Such point processes were analyzed also by Lewis and Shedler [24] in a model for transactions in a data base system. The case when $\lambda(K,t)$ is a stochastic process is discussed in my paper [7].

Now we assume that the organization of the work of operating system or the data management facility of information management system depends on the parameters λ /or μ /. This is the reason why we want to detect automatically at points A /or B/ any change in rates λ /or μ /. For such a system let us analyse the swapping policy in a virtual storage system /see Chow, Chin [14] /.

In virtual storage operating systems, control of the multiprogramming level is a major area of concern, particularly when some dynamic storage allocation scheme as the working-set strategy is used. However, total storage demand of the working sets changes dynamically with time and sometimes may reach some treshold, thus, causing an "overflow" event. An "underflow" event is caused when allocated storage falls below another treshold. At such events, swapping decisions can be made. If the margin of free storage is small then overflow and underflow events could occur quite frequently. On the other hand, if the margin is large, storage may not be utilized effectively /see Ryan and Coffman [27] /. In MVS, the paging rate is considered only in a swap-out decisions. Performance can be improved if the window size is varied to adapt to the changing needs of storage damand.

The swapping policy studied by Chow and Chiu [14] is of the "look-ahead" type, i.e., predictive values are used to evaluate the criteria for decision.

In practice such predictive values can be obtained from the empirical estimations of a queuing network model /see Fig. 2./.

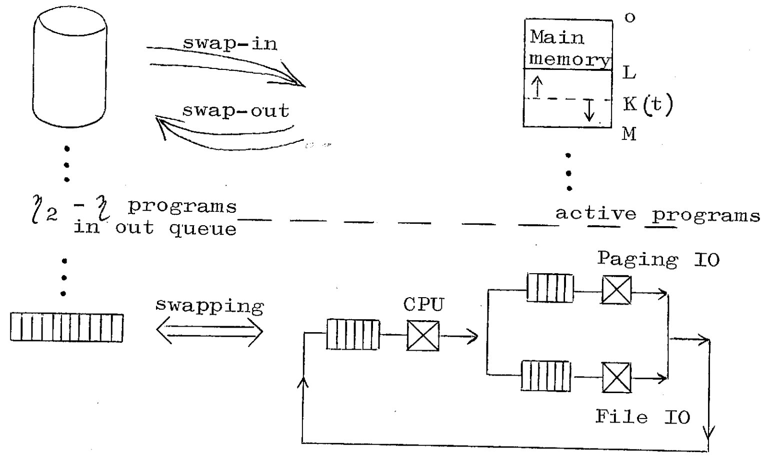

Figure 2.
Swapping and queuing model

The following notations are used:

η	degree of multiprogramming, $\eta_1 \le \eta \le \eta_2$,
T	working set window size, $T_1 \le T \le T_2$,
K (t)	number of allocated pages in main storage at time t,
L	underflow treshold of K (t),
M	overflow treshold of K (t) /capacity/,
U (η, T)	paging device utilization,
V (η, T)	CPU utilization.

The micro states $\{L \le K(t) \le M\}$ denote main storage utility which will cause, at the occurrence of overflows or underflows, a transition in macro states $\{(\eta, T, d)\}$, where d indicates the swapping decision.

Table 1

Present state: / η, T, d/. r_1 and r_2 are predetermined tresholds

Event	Criteria	Decision	New state
K(t) = M	$U(\eta, T-1) > r_1$	swapout	$(\eta-1, T, 1)$
	$U(\eta, T-1) \le r_1$	shrink	$(\eta, T-1, 2)$
K(t) = L	$U(\eta+1, T) \le r_2$	swapin	$(\eta+1, T, 3)$
	$U(\eta+1, T) > r_2$	enlarge	$(\eta, T+1, 4)$

If at some time t the number of allocated pages in main storage, K (t) reaches its upper or lower treshold the system may change ist degree of multiprogramming η or the window size T as it is summarized in Table 1.

A typical realization of K (t) is illustrated on Fig. 3., where the macro transitions are detected at times t_1, t_2,...

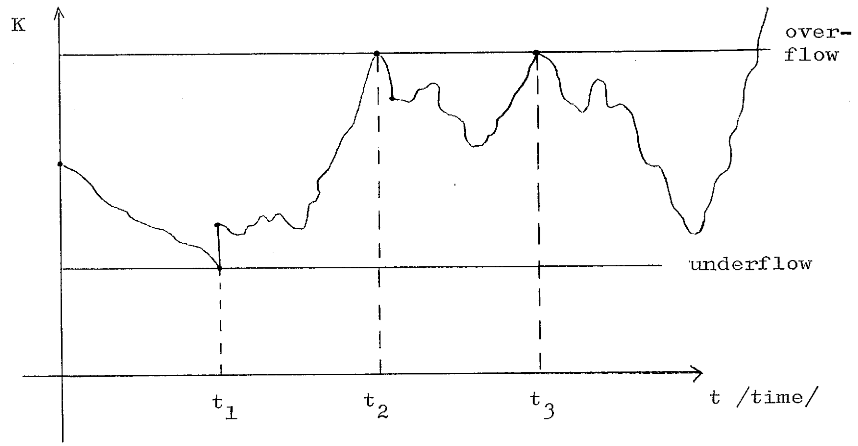

Figure 3.

For measures of performance effectivity we may use such values as overhead rate /the average amount of overhead per unit of time that the CPU is in problem state/, progress rate /proportion of time that the CPU is in the problem state/; the average degree of multiprogramming, the swapping rate, the paging rate, etc.

If we take one of these measures, e.g. the swapping rate, it exhibits significant changes /see Chow, Chiu [14]/. As the time points t_1, t_2, t_3,... forms approximately a Poisson process /see Cramer, Leadbetter [15] or Volkonskij, Rozanov [32]/we may assume that the swapping process forms also a Poisson type one, where the rate may be changed form λ_o to λ_1. The swapping policy /which means Table 1/ will be changed when a change form λ_o to λ_1 will be detected /a disorder occurs/. Analytical and numerical results for swapping rate in a queuing model are discussed in Chow, Chiu's [14] paper, where the detection /disorder/ problem was not treated.

b/ First we formulate the problem and its solution for Poisson process. The number events in the investigated process at A /or at B/ is denoted by N_t. The process N_t is assumed to be a Poisson process whose rate changes from λ_o to λ_1 ($\lambda_1 > \lambda_o > 0$/ at a certain time ξ . ξ is a random variable which is 0 with probability Π, and given that $\xi \neq 0$, exponentially distributed with parameter λ, i.e.

/2.1/ $$P\{\xi = 0\} = \Pi, \qquad P\{\xi < t \mid \xi > 0\} = 1 - e^{-\lambda t}.$$

We want to tell when ξ occured, from the observations of N_t. Thus the problem is to choose a stopping time τ/a random variable τ which depends only on the past i.e. $\{\tau < t\} \in \mathcal{G}\{N_s,\ s \le t\} = \mathcal{F}_t$, where $\mathcal{G}$ means the $\mathcal{G}$-algebra of events from the past/ so as to minimize the expected value of some cost function depending on the differences between τ and ξ. Let c, d positive constants and let $I_A(\omega)$ denote the characteristic /or indicator/ function of set A

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in \bar{A}. \end{cases}$$

We consider the following cost functions

/2.2/ $$\begin{aligned} s^1_\tau(\omega) &= d\cdot(\xi - \tau)\cdot I_{(\tau<\xi)} + c\cdot(\tau - \xi)\cdot I_{(\tau \ge \xi)}, \\ s^2_\tau(\omega) &= I_{(\tau < \xi - \varepsilon)} + c\cdot(\tau - \xi)\, I_{(\tau \ge \xi)}, \\ s^3_\tau(\omega) &= 1 - I_{(\xi - \varepsilon \le \tau \le \xi + \varepsilon)} \end{aligned}$$ /"hit or miss" cost/, where

$\varepsilon > 0$ constant.

The standard problem is to find the $\mathcal{F}_t$ - stopping time τ^i_o which minimizes $E s^i_\tau$, ($i = 1,2,3$, E means expectation).

To formulate the solution we introduce the process

/2.3/ $$\Pi_t = P(t \ge \xi \mid \mathcal{F}_t),$$

i.e. the aposteriori probability of random variable ξ after observing $\{N_s,\ s \le t\}$.

The evolution of Π_t can be written by the following stochastic differential equation /see Davis [16] formula /4.6//.

/2.4/ $$d\Pi_t = (\lambda_1 - \lambda_o)(\beta - \Pi_t)(1 - \Pi_t)dt + g(\Pi_{t-})\, dN_t$$

where

/2.5/ $$\beta = \frac{\lambda}{\lambda_1 - \lambda_o},$$

/2.6/ $$g(\Pi_{t-}) = \frac{(\lambda_1 - \lambda_o)\Pi_t - (1 - \Pi_{t-})}{\lambda_o(1 - \Pi_{t-}) + \lambda_1 \Pi_{t-}}$$

The following statement is true.

<u>Theorem.</u> If $\lambda_1 > \lambda_o$ and $k < \beta$, then

/2.7/ $$\tau^* = \inf\{t : \Pi_t \ge k\}$$

is optimal, where for s^1: $k = \frac{d}{d+c}$, for s^2: $k = \frac{\lambda'}{\lambda' + c}$, $\lambda' = \lambda e^{-\varepsilon\lambda}$.

The proof may be found in Davis' paper [16].

The cost function s^3 cannot be reduced to the standard form

/2.8/ $$s^k_\tau(\omega) = \int_0^\tau (\Pi_s - k)\,ds$$

because

/2.9/ $$E s^3_\tau = 1 + E(x_{\tau-\varepsilon} - x_{\tau+\varepsilon})$$

and $\tau - \varepsilon$ is not a stopping time in $\mathcal{F}_t$.

It can be shown that τ^* in general is not optimal in case $\lambda_1 > \lambda_o$ and $k > \beta$, or $\lambda_o > \lambda_1$ /see Galchuk, Rozovsky [18]/.

In this case the optimal time τ_o is

/2.10/ $$\tau_o = \inf\{t : \Pi_t \ge k_o\}$$

for some $k_o \in [k, 1]$. However, no simple way of finding the optimal k_o has yet been found. It involves the conditional distributions of the jump times of N_t, denoted by S_1, S_2, ...

<u>Remark 1.</u> Solution of equation (4) may be given numerically registrating step by step the realization of the process N_t.

<u>Remark 2.</u> In case when N_t is not a Poisson process the optimal stopping rule has not been found. The simple reason is that the observed process is not a Markovian one.

c/ In this part we discuss the diffusion approximation of the above mentioned problem. The Wiener process version of the disorder problem was raised up by Kolmogorov and was studid by Shiryayev. First of all we shall repeat a brief intuitive account and review of a diffusion approximation in the work of a computer system.

Diffusion approximation in the most general case for a single queue, which may be used for queuing networks and open networks, is given in an elagant paper of Gelenbe, Pujolle [21], where a detailed literature is also given. From another viewpoint we give also some exact results for diffusion approximation /see Arató [3], Arató, Knuth, Töke [9], Rét [26] / in utilization problem of CPU, a cyclic queue model and overhead time.

We recall that e.g. in the cyclic queue model $N(t)$, the number in queue at time t will be approximately a Wiener process with $EN(t) = (\lambda - \mu)t$, i.e. the drift is $\lambda - \mu$, and with variance $(\lambda^3 \cdot \sigma_a^2 + \mu^3 \sigma_s^2)\, t$, where the interarrival /service/ times have mean and variance λ^{-1} (μ^{-1}) and $\sigma_a^2 (\sigma_s^2)$ respectively /see Gelenbe, Pujolle [21]/.

In a similar way /see Arató [3]/ we can prove that the number of swappings in relatively great time intervals T_1, $2T_1$, $3T_1, \ldots$ is nearly normally distributed with parameters

$$E\left(N_{iT_1} - N_{(i-1)T_2}\right) = \lambda T_1,$$

$$D^2\left(N_{iT_1} - N_{(i-1)T_1}\right) = \lambda T_1.$$

This means that taking the process

$$w_{T_1} = N_{T_1},$$

$$w_{T_2} = N_{2T_1} - N_{T_1},$$

$$w_{T_i} = N_{iT_1} - N_{(i-1)\cdot T_1}, \ldots$$

we get in a new time axis a Wiener process with drift λT_1 and wariance λT_1 $(E\, w_{T_i} = \lambda T_1 \cdot i,\; D^2 w_{T_i} = \lambda T_1 \cdot i/.$

Now we can formulate the problem for a continuous time parameter Wiener process, assuming that the number of observation points $i \cdot T_1$ is very great. The problem can be stated roughly as follows. We observe a Wiener process whose drift changes from 0 to r /$r > 0$/ at a certain time ξ. It is a random variable, such that

/2.11/ $$P(\xi = 0) = \pi, \qquad P\{\xi \ge t \mid \xi > 0\} = e^{-\lambda t}, \quad t > 0,$$

where λ $(0 < \lambda < \infty)$ is a known constant and does not depend on π. The observed process ξ_t has the following differential form

/2.12/ $$d\xi_t = r \cdot I_{(t-\xi)} \cdot dt + \sigma\, dw(t),$$

where

$$\sigma > 0, \text{ /does not change/, } \quad I_t = \begin{cases} 1, & t \ge 0, \\ 0, & t < 0. \end{cases}$$

We consider the following cost function

/2.13/ $$s_\tau(\omega) = I_{(\tau < \xi)} + c \cdot (\tau - \xi) \cdot I_{(\tau \ge \xi)}, \quad c > 0,$$

and let

/2.14/ $$\rho(\pi) = \inf_\tau \left\{ P_\pi\{\tau < \xi\} + c\, E_\pi(\tau - \xi \mid \tau \ge \xi) \cdot P_\pi(\tau \ge \xi) \right\}.$$

The stopping time /markov moment/ τ^* is Bayesian if for all $0 \le \pi \le 1$:

/2.15/ $$P_\pi\{\tau^* < \xi\} + c \cdot E_\pi(\tau^* - \xi \mid \tau^* \ge \xi)\, P_\pi(\tau^* \ge \xi) = \rho(\pi).$$

Let

$$\pi_t = P_\pi\{\xi \le t \mid \mathcal{F}_t\}$$

the posterior probability distribution of random variable ξ. The evolution of π_t can be written by the following way /see Shiryayev [29] /4.151//:

/2.16/ $$d\pi_t = \lambda(1 - \pi_t)\, dt + \frac{r}{\sigma} \pi_t (1 - \pi_t)\, d\bar{w}(t)$$

where $w(t)$ is a Wiener process, or /Shiryayev [29], /4.160//:

/2.17/ $$d\pi_t = (1 - \pi_t)\left(\lambda - \frac{r^2}{\sigma^2} \pi_t^2\right) dt + \frac{r}{\sigma^2} \pi_t (1 - \pi_t)\, d\xi_t.$$

For the Wiener process the following statement is true.

Theorem 2. The Bayesian stopping time τ^* exists and

/2.18/ $$\tau^* = \inf\{t \ge 0 : \pi_t \ge A^*\}.$$

A^* is the solution of the equation:

/2.19/ $$C^{-1} = \int_0^{A^*} e^{-\Lambda[H(A^*) - H(x)]} \cdot \frac{dx}{x(1-x)^2}$$

where

$$C = c\left(\frac{2\sigma^2}{r^2}\right), \quad \Lambda = \lambda \frac{2\sigma^2}{r^2} \quad \text{and} \quad H(x) = \ln \frac{x}{1-x} - \frac{1}{x}.$$

The proof may be found in Shiryayev's book [29].

Remark 1. As a consequence of theorem 2. we get that among the stopping times τ, for which $P_\pi(\tau < \xi) \le \alpha, (0 \le \alpha < 1)$, the optimal $\tilde{\tau}$, where

$$/2.20/ \qquad E_{\pi}(\tilde{\tau} - \xi \mid \tilde{\tau} \geq \xi) \leq E_{\pi}(\tau - \xi \mid \tau \geq \xi),$$

is the following

$$/2.21/ \qquad \tilde{\tau} = \inf \{ t \geq o : \pi_t \geq 1 - \alpha \}.$$

Remark 2. The solution /or development/ of equation /17/ may be given only numerically by the registration of the process ξ_t.

3. INDEPENDENT REFERENCE STRING MODELS

a/ There are two main models for a reference /or request/ string $\zeta_1, \zeta_2, \ldots$. The first one is the independent reference model, when the reference string ζ_t means a sequence of independent indentically distributed random variables. The second one is the LRU stack model where the distance string D_t $(t = 1,2,\ldots)$, e.g. sequences of stack distances for least recently used /LRU/ replacement, is a sequence of independent, identically distributed random variables. D_t is the total number of distinct references since the last reference to ζ_t. Lewis and Shedler [23] statistically proved that both of these models are not exactly adequate in paged computer systems. In this part we use the independent reference string model. Let $\zeta_1, \zeta_2, \ldots, \zeta_t, \ldots$ denote the reference string, then by our assumption this sequence of random variables is independent, identically distributed

$$P\{\zeta_t = i\} = p_i, \; i = 1,2,\ldots, n.$$

At first we assume that the probabilities p_i are known and for simplicity $p_1 \geq p_2 \geq \ldots \geq p_n > 0$. $\sum_1^n p_i = 1$.

Let d_t mean the subset of indices /d_t consist of n-m elements/. We assume that decision d_t, depends only on the initial decision d_o and the observed reference string $\zeta_1, \ldots, \zeta_{t'}$, $t' \in [t, N)$. 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 $[t,N)$. In our first model, case A, the loss function has the following form:

$$/3.1/ \qquad X_t^{d_{t-1}} = \begin{cases} 1, & \text{if } \zeta_t \in d_t, \\ 0, & \text{other wise.} \end{cases}$$

In the other case investigated by us, case B, the loss function has the following form:

$$/3.2/ \qquad X^{d_t, d_{t-1}} = |d_t \setminus d_{t-1}|,$$

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

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

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

$$\left(v(N) = E\left[\sum_{t=1}^{N} X_t^{d_t, d_{t-1}}\right]\right),$$

in case A /resp. B/.
Let us denote the conditional expectation under a given realization $y_1, \ldots, y_t$ of $\zeta_1, \ldots, \zeta_t$ by $E_{\{y_1, \ldots, y_t\}}$ and define the families of conditional risk functions

$$/3.3/ \qquad v(y_1, \ldots, y_t, N-t) = \min_{\{d_t, \ldots, d_{N-1}\} \in D_{t,N}} E_{\{y_1, \ldots, y_t\}} \sum_{\tau=t+1}^{N} X_\tau^{d_{\tau-1}}$$

and

$$/3.4/ \qquad v(y_1, \ldots, y_t, d_t, N-t) = \min_{\{d_{t+1}, \ldots, d_{N-1}\} \in D_{t+1,N}} E_{\{y_1, \ldots, y_t\}} \sum_{\tau=t+1}^{N} X_\tau^{d_\tau, d_{\tau+1}}$$

in case A and B respectively. Families /3/ and /4/ statisfy the Bellman equations

$$/3.5/ \qquad v(y_1, \ldots, y_{t-1}, N-t+1) = \min_{d_{t-1}} E_{\{y_1, \ldots, y_{t-1}\}}\left[X_t^{d_{t-1}} + v(y_1, \ldots, y_{t-1}, \zeta_t, N-t)\right],$$

$$/3.6/ \qquad v(y_1, \ldots, y_{t-1}, d_{t-1}, N-t+1) = \min_{d_t} E_{\{y_1, \ldots, y_{t-1}\}}\left[X_t^{d_t, d_{t-1}} + v(y_1, \ldots, y_{t-1}, \zeta_t, d_t, N-t)\right].$$

Solving recursively systems of equations /5/ and /6/ we can find the optimal strategies. In case A $v(y_1, \ldots, y_t, N-1)$ does not depend on d_{t-1} therefore it is sufficient to minimize for every t the conditional expectation

$$E_{\{y_1,\dots,y_{t-1}\}}\left[X_t^{d_{t-1}}\right].$$

The same statement, with a slight modification, is true in case B.

As

/3.7/ $$\min_{d_{t-1}} E_{\{y_1,\dots,y_{t-1}\}}\left[X_t^{d_{t-1}}\right] = \left(p_n+p_{n-1}+\dots+p_{m+1}\right)$$

and

/3.8/ $$\min_{d_t} E_{\{y_1,\dots,y_{t-1}\}}\left[X_t^{d_t,d_{t-1}}\right] =$$

$$= \left(p_n + p_{n-1} + \dots + p_{m+2} + p_\ell\right),$$

where $1 \le \ell \le m + 1$.

We have thus proved the following result, which generalizes a theorem of A. Lew [27].

Theorem 3. Assume that the reference string ζ_t is an independent sequence with stationary probabilities. Then the policy of taking that index with the least probability of reference is optimal.

b/ The reference string $\zeta_1, \zeta_2, \dots, \zeta_t$ from probabilistic point of view forms a sequence of independent identically distributed random variables; the common probability distribution

$$P_{i,w} = P_w\left(\zeta_t = i\right)$$

of the random variables ζ_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 $1, \dots, n$; $w(i)$ denotes the one to one mapping of set $\{1,\dots,n\}$ realized by w. There is given a fixed decreasing sequence $p_1 > \dots > p_n > 0$ of probabilities $p_1 + \dots + p_n = 1$ and

$$\left\{P_{i,w}\right\} = \left\{P_w\left(\zeta_t = i\right)\right\} = \left\{p_{w(i)}\right\}.$$

Following the Bayesian approach in decision theory we assume that w itself is a random variable. As we have no preliminary information about the distribution $P_w(\zeta_t = i)$ we assume that the prior distribution of parameter w is the uniform one. The optimality of LFU strategy is a consequence of the following lemma /see Benczúr, Krámli, Pergel [11]/.

Lemma 1. If the prior probability distribution of the parameter w is the uniform one and the frequency f_i of page i in the string $\{y_1,\dots,y_{t-1}\}$ is less than that of page j, f_j, then

/3.7/ $$P\left(\zeta_t = i \mid y_1,\dots,y_{t-1}\right) < P\left(\zeta_t = j \mid y_1,\dots,y_{t-1}\right),$$

i.e. the order of posterior probabilities of indices after observing the string $y_1,\dots,y_{t-1}$ is the same as the order of their frequencies in this string.

Proof. Let i and j two indices such that $f_i < f_j$. If w_1 and w_2 are two permutations with the properties

/i/ $w_1(i) = w_2(j)$,

/ii/ $w_2(i) = w_1(j)$,

/iii/ $w_1(i) < w_1(j) \Leftrightarrow w_2(j) < w_2(i)$,

/iv/ $w_1(k) = w_2(k)$ for $k \neq i,j$,

then, using Bayes' theorem

$$P\left\{w_1 \mid y_1,\dots,y_t\right\} = \frac{\prod_{k=1}^{n} p_{w_1(k)}^{f_k}}{\sum_{w\in W} \prod_{i=1}^{n} p_{w(i)}^{f_i}}$$

and

$$P\left\{w_2 \mid y_2,\dots,y_t\right\} = \frac{\prod_{k=1}^{n} p_{w_2(k)}^{f_k}}{\sum_{w\in W} \prod_{i=1}^{n} p_{w(i)}^{f_i}}$$

we get

$$P\left\{w_1 \mid y_1,\dots,y_t\right\} < P\left\{w_2 \mid y_1,\dots,y_t\right\}.$$

Summing these probabilities we get the required result.

Using Lemma 1 and the uniformity of the prior distribution of parameter w we get the following statement.

Theorem 4./Separation principle./ The least frequently used strategy minimizes the expected loss $E\left[\sum_{t=1}^{N} X_t^{d_{t-1}}\right]$ in case A, where d_0 is arbitrary, and the initial distribution ξ of random variable w is uniform.

In case B we must argue more carefully. First we prove that the optimal strategies are among the demand paging algorithms, i.e. among the algorithms statisfying the conditions

$$d_t = d_{t-1}, \quad \text{if} \quad \zeta_t \bar{\in} d_{t-1},$$

$$d_{t-1} \setminus d_t = \{\zeta_t\}, \text{if} \quad \zeta_t \in d_{t-1}.$$

The optimality of the LFU strategy in case B follows from Theorem 5.

Theorem 5. 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 is less, than the frequency f_j of the page j, then

$$V(y_1,\dots,y_t,\ d_t,\ N-t) < V(y_1,\dots,y_t,\ d_t',\ N-t).$$

The proof can be carried out by induction for $\Theta = N-t$. The assertion of Theorem 5 for $\Theta = 1$ is an obvious consequence of the observation used in case A /Lemma 1./. The proof of the induction step - the comparison of conditional risk functions

$$V(y_1,\dots,y_t,\ d_t,\ N-t) = E_{\{y_s,\ s\leq t\}}\left[\sum_{s=t}^{N} X_s^{d_s,\ d_{s-1}}\right]$$

and

$$V(y_2,\dots,y_t,\ d_t',\ N-t)$$

for $t < N-1$ is not so simple as in case A. Here we use essentially the fact that $V(y_1, y_2,\dots,y_t,\ d_t,\ N-t)$ depend only on the frequencies of the pages in the string $\{y_1, y_2,\dots,y_t\}$ and on d_t. A detailed proof for page replacement algorithms is given in the paper Benczúr, Krámli, Pergel [11].

c/ The optimality of the LFU strategy without Bayesian assumption we discuss too /see Arató, Benczúr, Krámli [8]/.

Definition 1: A sequence $\{\delta_0,\dots,\delta_{N-1}\}$ of probability distributions on the space Δ of all possible subsets d consisting of /n-m/ elements of the set $\{1,\dots,n\}$ is called randomized sequential decision procedure if and only if for every $0 \leq t \leq N-1$ the probability distribution δ_t depends only on the prior distribution ξ, and the reference string $\{\zeta_1,\dots,\zeta_t\}$.

Definition 2: A randomized sequential decision procedure $\{\delta_0, \delta_1, \dots, \delta_{N-1}\}$ is called symmetric if and only if for every moment $0 \leq t \leq N$ permutation $w \in W$ and realization $\{y_1, \dots, y_t\}$ the following holds

$$w(\delta_t(\xi,\ y_1,\dots,y_t)) = \delta(\xi,\ w(y_1),\dots,w(y_t)).$$

Let us define the "action" $w(\delta)$ of a permutation $w \in W$ on a δ distribution by relation $P_{w(\delta)}(d) = P_\delta(w^{-1}(d))$. The following statement is true.

Lemma 2. For both forms of the loss function /cases A and B/ among the randomized decision procedures the corresponding LFU strategies are optimal, i.e. the procedures $\{\delta_0,\dots,\delta_{N-1}\}$ for which the measure δ_t is concentrated on the subsets d of the set $\{1,2,\dots,n\}$ $(\{1,2,\dots,n\} \setminus \{\zeta(t)\})$ in case A(B) consisting of the least frequently used indices in the string $\{\zeta_1,\dots,\zeta_{t-1}\}$.

Notice that in the case of strictly different frequencies δ_t is concentrated on a unique subset d.

Proof. The assertion follows from the fact that in relation (7) stands the strict inequality $<$, if $f_i < f_j$.

We had to extend the sequential decision procedure to the randomized case too if we wanted to preserve the symmetry for the LFU strategies. A nonrandomized LFU strategy cannot be symmetric.

Theorem 6. If the prior distribution ξ is concentrated on a unique permutation $w \in W$, then for both forms of the loss-function among the symmetric randomized sequential decision procedures the LFU strategies are optimal. (The proof may be found in Arató, Benczúr, Krámli [8]).

Remark 1. The theorem states that the LFU strategies are optimal in the non-Bayesian case too.

Remark 2. In a problem of dynamic file assignement let $Y_i(t)$ /$i = 1,2,\dots, n$; $t = 1,2,\dots$/ take the value 1 or 0 according to whe ther the file is located in memory of i-th computer at time t. Further let $\zeta(t) = \{\zeta_1(t),\dots,\zeta_n(t)\}$ the request string of the file by the n computers. Then

$$X_t^{d_{t-1}} = \sum_{\substack{i,j\\ i\neq j}} \zeta_j(t)\, Y_i(t) = \sum_{j\neq d_{t-1}} \zeta_j(t)$$

and

$$P\{\zeta_j(t) = 1\} = p_j.$$

Remark 3. In demand page replacement algorithms in case A. a page from the second level is delivered to the m + 1 - th place /$m < n$ pages are in the first level/ and after delivering the content of it a page must be removed to the second level. Then, if d_t means the subset of pages being absent of the central memory after moment t,

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

In case B each change of a page increases the cost at moment t by 1 and ζ_t must be stored in the central memory. A page on the first level must be replaced by the requested one.

The loss function has the form

$$X^{d_t, d_{t-1}} = |d_t \setminus d_{t-1}|,$$

where $|\cdot|$ denotes the number of elements of a set.

References

[1] AHO A.V., DENNING P.J., ULLMANN J.D. /1971/ Principles of optimal page replacement, Journal A.C.M. 18/1, 80-93.

[2] ARATÓ, M. /1973/ On computing technology for mathematical statistics and stochastic processes and its application for evaluating computer system performance, Proc. Computer Science Conference, Hungary, 231-237.

[3] /1975/ Diffusion approximation for multiprogrammed computer systems, Comp. and Maths. with Appls. 1, 315-326.

[4] /1976/ Számitógépek hierarchikus laptárolási eljárásainak optimalizálásáról, MTA SZTAKI Közlemények v. 16, 7-23, /in Hungarian/.

[5] /1976/ Statistical sequential methods in performance evaluation of computer system, 2nd Internat. workshop on modelling and performance evaluation of computer systems, Stresa-Italy, 1-10.

[6] /1976/ A note on optimal performance of page storage, Acta Cybernetica, Tom 3, fasc. 1, 25-30.

[7] /1977/ Statistical Sequential Decision Methods in Case of Independent Strings, Second Hungarian Computer Sci. Conference Budapest, 33-57.

[8] ARATÓ M., BENCZÚR A., KRÁMLI A. /1977/ On the solution of optimal performance of page storage hierarchies with independent reference string. Banach Center Publications, Vol. 6., Mathematical Statistics, /in print/.

[9] ARATÓ M., KNUTH E., TŐKE P. /1974/ On stochastic control of a multiprogrammed computer based on a probability model, Preprints of the Stochastic Control Symposium /IFAC, Budapest/ 305-313.

[10] ASZTALOS D. /1977/ Performance measurement of MULTIJOB operating System, In this volume

[11] BENCZÚR A., KRÁMLI A., PERGEL J. /1976/ On the Bayesian approach to optimal performance of page storage hierarchies, Acta Cybernetica, v. 3.

[12] BENCZÚR A., KRÁMLI A. /1977/ A probabilistic model for the comparison of different memory handling systems, Second Hungarian Computer Science Conference, Budapest, 178-194.

[13] BÉLÁDY L. /1966/ A study of replacement algorithms for a virtual storage computer, IBM Syst. Journ. 5, 78-101.

[14] CHOW, W.M., CHIU W.W. /1977/ Analysis of swapping policies in virtual storage systems, IEEE Trans. on software Eng. Vol. Se-3, 150-156.

[15] CRAMER H., LEADBETTER M. /1967/ Stationary and related stochastic processes, J. Wiley, London.

[16] DAVIS, M.H.A. /1976/ A Note on the Poisson Disorder Problem Banach Center Publications, Vol. 1, 65-72, Mathematical Control Theory.

[17] DE GROOT, M.H. /1970/ Optimal statistical decisions, Mc Graw-Hill, N.Y.

[18] L.I. GALCHUK, B.L. ROZOVSKY /1971/ The disorder problem for a Poisson process, Theory of Prob. and Appl. 16, 729-734.

[19] GELENBE, E. /1973/ A unified approach to the evaluation of a replacement algorithms, IEEE Trans. Computers, v. C-22, 611-617.

[20] GELENBE, E., MUNTZ R.R. /1976/ Probabilistic Models of Computer Systems-Part I. /Exact Results/, Acta Informatica 7, 35-60.

[21] GELENBE, E., PUJOLLE, G. /1976/ The Behaviour of a Single Queue in General Queuing Network, Acta Informatica 7, 123-126.

[22] LEW, A. /1976/ Optimal Control of demand - paging systems, Information Sciences, 10, 319-330.

[23] LEWIS, P., SHEDLER, G. /1973/ Empirically derived micromodels for sequences of page exceptions, IBM J. Res. Developm. March, 86-100.

[24] /1975/ Statistical analysis of transaction processing in a data base system, Sept. R.J. 1629. IBM Research.

[25] LIPCER, R., SHIRYAYEV, A. /1974/ Statistics of Stochastic Processes, Izd. Nauka Moscow /In Russian/.

[26] RÉT, M. /1977/ On the use of diffusion approximations for the cyclic queue models, MTA SZTAKI Közlemények 18, 99-104.

[27] T.A. RYAN; E.G. COFFMAN /1974/ A problem in multiprogrammed storage allocation, IEEE Trans. Comput. vol. c-23, N°11.

[28] SEGALL, A. /1976/ Dynamic file assignment in a computer network, IEEE Trans. on Automatic Control, Vol. AC-21, N°2, 161-173.

[29] SHIRYAYEV, A. /1969/ Statistical sequential analysis, Izd. Nauka, Moscow /in Russian/.

[30] SHOHAT, Y., STRAUSS, J.C. /1976/ An analytic model of dispatching algorithms. Modelling and Performance Evaluation of Computer Systems, 335-362.

[31] TŐKE, P. /1977/ Issledovanie sloznih algoritmov massovogo obsluzivanija, Second Hungarian Computer Science Conference, Budapest, 871-886.

[32] VOLKONSZKIJ, V.A., ROZANOV, J.A. /1959, 1961/ Some limit theorems for random functions, I and II, Theory of Prob. and Appl. 4, 186-207, and 6, 202-215.

[33] WONHAM, W.M. /1968/ On the separation theorem of stochastic control, SIAM Journ. Control 6, 312-326.