Alkalmazott Matematikai Lapok 3 *(1977) 13—25*

SZEKVENCIÁLIS STATISZTIKAI DÖNTÉSI MÓDSZEREK FÜGGETLEN HIVATKOZÁSOK ESETÉBEN[1]

ARATÓ MÁTYÁS

Budapest

Az elmúlt években jelentősen megnőtt az érdeklődés a számítógéprendszerek modellezésének és elemzésének problémái iránt, beleértve az interaktív számítógép-hálózatokat is. A cikkek legtöbbjében vizsgált statikus elemzésnél feltették, hogy a rendszer minden paramétere (igénybevétel mértéke, tárolási költség stb.) előre ismert, és hogy a tervezés ezek átlagértékén alapul. Amikor a rendszer paraméterei időben változnak vagy ismeretlenek, a dinamikus kezelés nemcsak lényeges javítását eredményezi a működésnek, de léteznek speciális esetek, mikor viszonylag egyszerű megoldásokat kapunk.

Ennek a cikknek a célja, hogy egzakt formájú optimális politikát találjunk a file-ok dinamikus elhelyezésére, a lapcserére és a tranzakciók sebességparaméterének becslésére. Ebben a dolgozatban viszonylag egyszerű modelleket tekintünk, hogy az ilyen jellegű problémák megoldásának fő módszereit bemutassuk. A statisztikai szekvenciális döntéselmélet jól ismert módszereit használjuk. Ilyen rendszerek elemzésében az alkalmazásra irányuló közlemények mellett (lásd korábbi munkáinkat: Arató [3], [4] vagy Arató, Benczúr, Krámli [5], Benczúr, Krámli, Pergel [8]) a szekvenciális módszerek (vagy optimális szabályozás) elméletében elvileg új eredményeink vannak. Az első fejezet bemutat egy alapmodellt, amelyet számítógép-hálózatokban file kijelölésre használunk független hivatkozási string esetén. A problémával A. Segall [19] cikkében általánosan foglalkozott, de nem kapott egyszerű megoldásokat, mivel modellje bonyolult volt. A második fejezet tartalmazza azokat a korábbi eredményeinket, amelyek a laptárolásra vonatkoztak, amikor a hivatkozási string egy független sorozat. Új tételt adunk meg ismert valószínűségek esetén, amely A. Lew [16] eredményének általánosítása. Ezen kívül ez a tétel megmutatja az összefüggést a szekvenciális módszer és a dinamikus programozás között. A harmadik fejezetben egy adatbázis rendszer tranzakciói számára adunk egy új modellt, amelynek statisztikai vizsgálatát és elemzését Lewis és Shedler [18] végezték el. Ilyen típusú modellt korábbi cikkeinkben használtunk. (Arató [2], Arató, Knuth és Tőke [6], Knuth [15]), de statisztikai elemzés és adatgyűjtés nélkül.

1. Dinamikus file elhelyezés független hivatkozási stringekkel

Ebben a részben egy dinamikus file elhelyezési problémával foglalkozunk, egy n-számítógépes hálózatban, amikor a file-igény mérték sebességei ismertek, és később, amikor azok nem ismertek. Feltesszük, hogy a file csak az egyik számítógép memóriájában tárolható, de elérhető a többi által is. Feltesszük, hogy az idő diszkrét és a számítógépek teljes kapcsolatban vannak.

Az $Y_i(t)$ $(i=1, 2, \ldots, n;\ t=1, 2, \ldots)$ változók az 1 vagy a 0 értéket vehetik fel, attól függően, hogy a file az i-edik számítógép memóriájában van a t időben, vagy egy másiknál, azaz ha

$$Y_i(t) = 1, \quad \text{akkor} \quad Y_1(t) = \ldots = Y_{i-1}(t) = Y_{i+1}(t) = \ldots = Y_n(t) = 0.$$

[1] A II. *Magyar Számítástudományi Konferencián* elhangzott előadás. (Preprint I. kötet, 33—57. old.)

Jelöljék $\eta_j(t)$ $(j=1, 2, \ldots, n; t=1, 2, \ldots)$ valószínűségi változók sorozatait, amelyek leírják az n számítógép általi file igényeket

$$\eta_j(t) = \begin{cases} 1, & \text{ha van kérés a } t \text{ időben a } j\text{-edik számítógépnél,} \\ 0, & \text{egyébként.} \end{cases}$$

Feltesszük, hogy az $\eta_j(t)$ változók teljesen független *Bernoulli sorozatokat* alkotnak és ugyanarra a j-re azonos eloszlású valószínűségi változók

$$P\{\eta_j(t) = 1\} = p_j, \quad j = 1, 2, \ldots, n; \quad t = 1, 2, \ldots. \tag{1.1}$$

Az egyszerűség kedvéért legyen $p_1 \geqq p_2 \geqq \ldots \geqq p_n > 0$. Feltesszük továbbá, hogy a műveleti költség 0, amikor a file abban a gépben van, amely kéri, és az átviteli költség az egyik számítógépből a másikra való átmenetként 1. Ez esetben a teljes költség várható értéke egy $t=1, 2, \ldots, N$ periódusra

$$C = \mathbf{E}\left\{\sum_{t=1}^{N} \sum_{\substack{i,j \\ i \neq j}} \eta_j(t) Y_i(t)\right\}. \tag{1.2}$$

Definiáljuk a d_t döntéseket a t időben $(t=0, 1, 2, \ldots)$ a következőképpen: ha a szóban forgó $d_t = i$, akkor a file-t az i-edik számítógép tárolja a $t+1$ időben (azaz t után). Ez azt jelenti, hogy

$$Y_i(t) = \begin{cases} 1, & \text{ha } d_{t-1} = i, \\ 0, & \text{egyébként.} \end{cases}$$

Most feltesszük, hogy a d_t döntés az előző megfigyelések $\{\eta_1(s), \ldots, \eta_n(s), d_s, s \leqq t-1\}$ függvénye. A feladat C minimalizálása (1.2)-ben. Az esetben, amikor a hivatkozási valószínűségek $p_1 \geqq \ldots \geqq p_n > 0$ ismertek, a következő állítás igaz.

1.1. TÉTEL. *Tegyük fel, hogy az $\eta_j(t)$ hivatkozási folyamatok független Bernoulli folyamatok, $p_1 \geqq \ldots \geqq p_n > 0$, (lásd (1.1)), akkor a $d_t \equiv 1$ döntési politika optimális, azaz ez minimalizálja (1.2)-t.*

Bizonyítás. Jelölje $y_j(s)$ $(s \leqq t)$ $\eta_j(t)$ egy realizációját és $\mathbf{E}\{y_j(1), \ldots, y_i(t)\}$ a feltételes várható értéket ezen feltétel mellett. Definíció szerint legyen

$$v(y_j(1), \ldots, y_j(t), \ j = 1, 2, \ldots, n, \ N-t) = \tag{1.3}$$

$$\min_{\{d_t, \ldots, d_{N-1}\} \in D_{t,N}} \mathbf{E}\{y_j(1), \ldots, y_j(t), \ j = 1, 2, \ldots, n\} \cdot \left[\sum_{\tau=t+1}^{N} \sum_{\substack{i,j \\ i \neq j}} \eta_j(\tau) Y_i(\tau)\right],$$

ahol $D_{t,N}$ jelöli az összes lehetséges *Markov szekvenciális döntési eljárások* halmazát. A $v(y_j(1), \ldots, y_j(t); j=1, 2, \ldots, n; N-t)$ függvény kielégíti az ún. *Bellmann egyenletet:*

$$v(y_j(1), \ldots, y_j(t-1); \ j = 1, 2, \ldots, n; \ N-t+1) = \tag{1.4}$$

$$= \min_{d_{t-1}} \mathbf{E}\{y_j(1), \ldots, y_j(t-1), \ j = 1, 2, \ldots, n\} \cdot$$

$$\cdot \left[\sum_{\substack{i,j \\ i \neq j}} \eta_j(t) Y_i(t) + v(y_j(1), \ldots, y_j(t-1), \eta_j(t); \ j = 1, 2, \ldots, n; \ N-t)\right].$$

Mivel $v(y_i(1), ..., y_i(t); j=1, ..., n; N-t)$ nem függ d_{t-1}-től, minden t lépésnél a következő kifejezést kell minimalizálnunk

$$\min_{d_{t-1}} \mathbf{E}\{y_j(1), ..., y_j(t-1), j = 1, ..., n\} \sum_{\substack{i,j \\ i \neq j}} \eta_j(t) Y_i(t) =$$

$$= \min_{d_{t-1}} \mathbf{E}\{y_j(1), ..., y_j(t-1)\}[X_t^{(d_{t-1})}] =$$

$$= \mathbf{E}\left[\sum_{j=2}^{n} \eta_j(t)\right] = (p_2+p_3+...+p_n),$$

ahol $X_t^{d_{t-1}} = \sum_{j \neq d_{t-1}} \eta_j(t)$.

Ez bizonyítja a tételt. A továbbiakban feltesszük, hogy a számítógépek hivatkozási valószínűségei nem ismertek, és ezért használjuk a *Bayes-féle eljárást*. Jelölje w azt a valószínűségi változót, melynek értelmezési tartománya az $1, 2, ..., n$ természetes számok összes permutációjának halmaza; $w(i)$ jelöli a $\{1, 2, ..., n\}$ halmaz w-vel történő egy-egy értelmű leképezését. A *Bernoulli sorozatok* eloszlásai a következőképpen adottak:

$$P_w\{\eta_j(t) = 1\} = \{P_{j,w}\}_{j=\overline{1,n}} = \{p_{w(i)}\}. \tag{1.5}$$

Feltesszük, hogy a w paraméter apriori eloszlása egyenletes, mivel nincs erről más előzetes információnk:

$$P\{w = k\} = \frac{1}{n!}. \tag{1.6}$$

Ugyanazt a megjegyzést téve, mint fent, minimalizálnunk kell az

$$\mathbf{E}\left[\sum_{t=1}^{N} \sum_{\substack{i,j \\ i \neq j}} \eta_j(t) Y_i(t)\right] = \mathbf{E}\left[\sum_{t=1}^{N} X_t^{(d_{t-1})}\right] \tag{1.2'}$$

kifejezést, ahol

$$X_t^{(d_{t-1})} = \eta_1(t)+...+\eta_{d_{t-1}-1}(t)+\eta_{d_{t-1}+1}(t)+...+\eta_n(t) = \sum_{j \neq d_{t-1}} \eta_j(t).$$

Ez esetben

$$P\{w = k | \eta_j(1), ..., \eta_j(t), j = 1, 2, ..., n\}$$

függ w apriori eloszlásától és a j számítógépben a file f_j hivatkozási gyakoriságától ($j=1, ..., n$). Ugyanúgy, mint két számítógépes rendszer esetében (lásd ARATÓ [3]) a *Bayes tétel* használatával bebizonyítjuk a következő lemmát (lásd még BENCZÚR, KRÁMLI, PERGEL [8]).

1.1. LEMMA. Ha a w paraméter apriori eloszlása egyenletes és a file f_j hivatkozási gyakorisága az $\{\eta_j(1), ..., \eta_j(t-1)\}$ stringben kisebb, mint f_i az $\{\eta_i(1), ..., \eta_i(t-1)\}$ stringben, akkor

$$P\{\eta_j(t) = 1 | \eta_k(1), ..., \eta_k(t-1), \quad k = 1, 2, ..., n\} <$$
$$P\{\eta_i(t) = 1 | \eta_k(1), ..., \eta_k(t-1), \quad k = 1, 2, ..., n\}.$$

Megjegyzés. A lemma azt állítja, hogy a file hivatkozások aposteriori valószínűségeinek sorrendje az $\{\eta_j(1), \ldots, \eta_j(t-1),\ j=1, 2, \ldots, n\}$ string megfigyelése után ugyanaz, mint a gyakoriságuk sorrendje a stringjükben. Most már kimondhatjuk a probléma optimalitási kritériumát. Az ún. szeparációs tétel egzakt megoldását kapjuk, mely hasonló WONHAM lineáris *Gauss folyamatokra* vonatkozó szeparációs tételéhez (lásd [19]). Ezt a szeparációs tételt dinamikus file kijelölésre SEGALL kimondta, de nem oldotta meg konkrét esetben.

1.2. TÉTEL. Legyen $p_1 \geqq p_2 \geqq \ldots \geqq p_n$ és N rögzített. Az $\eta_j(t)$ $(j=1, 2, \ldots, n,$ $t=1, 2, \ldots)$ hivatkozási stringek független *Bernoulli sorozatok* (1.5) hivatkozási valószínűségekkel, ahol w egyenletes eloszlású (1.6). Az optimális szekvenciális döntési eljárás, amelyik minimalizálja az átvitelek várható számát (1.2'), a file-t bármelyik időpontban ahhoz a számítógéphez helyezi, amelyik többször kérte, mint a többi.

Bizonyítás. A bizonyítás újból a *Bellmann egyenleten* alapul.

Legyen

$$V\big(\eta_j(s),\ j = 1, 2, \ldots, n,\ s \leqq t,\ N-t,\ \delta^{[t,N]} = (d_t, \ldots, d_{N-1})\big) =$$

$$= \mathbf{E}\{\eta_j(s),\ j = 1, 2, \ldots, n,\ s \leqq t\} \left[\sum_{s=t}^{N} X_s^{(d_s-1)}\right],$$

$$V\big(\eta_j(s),\ j = 1, 2, \ldots, n,\ s \leqq t,\ N-t\big) =$$

$$= \inf_{\delta^{[t,N]}} V\big(\eta_i(s),\ j = 1, \ldots, n,\ s \leqq t,\ N-t,\ \delta^{[t,N]}\big).$$

Egyszerű számítással kapjuk, hogy

$$V\big(\eta_j(s),\ j = 1, \ldots, n,\ s \leqq t-1,\ N-t+1\big) =$$

$$= \inf_{d_{t-1}} \mathbf{E}\{\eta_j(s),\ j = 1, \ldots, n,\ s \leqq t-1\} \cdot$$

$$\cdot \big[X_t^{(d_t-1)} + V\big(\eta_j(s),\ j = 1, \ldots, n,\ s \leqq t,\ N-t\big)\big],$$

ahol $V\big(\eta_j(s), j=1, \ldots, n, s \leqq t, N-t\big)$ nem függ d_{t-1}-től. Ezután az 1.1. lemmát használva megkapjuk az állítást.

2. Laphelyettesítési algoritmus

Ebben a részben a helyettesítés *(demand-paging)* problémájával foglalkozunk. Ez a feladat azt a döntést jelenti, hogy egy lapolási hiba előfordulása esetén melyik lap legyen kicserélve az ebben az időben a központi memóriában levő lapok közül. Az általában elfogadott optimalitási feltétel az, hogy úgy kell cserélni a lapokat, hogy minimalizáljuk a lapolási hibák (vagy cserék) számát. Ez a politika általában megvalósíthatatlan, mivel a jövőbeli viselkedés pontos ismeretét követeli meg. Az optimális politika valószínűségelméleti kiterjesztése azt jelenti, hogy a hivatkozási string adott eloszlására minimalizáljuk a lapolási hibák várható számát. A laphivatkozási string egy valószínűségeloszlási osztályára optimális helyettesítési algoritmus nem szükségképpen optimális más osztályokra.

Két fő modell ismeretes, az $\eta_1, \eta_2 \ldots$ hivatkozási stringekre. Az első a független hivatkozási modell, amikor az η_t hivatkozási string független azonos eloszlású valószínűségi változók sorozatát jelenti. A második az *LRU stack modell*, ahol a D_t ($t=1, 2, \ldots$) távolsági stringben — a legrégebben használt lap kerül cserélésre *(LRU modell)* — a változók független, azonos eloszlású valószínűségi változókat jelentenek. D_t azon különböző lapok száma, melyekre hivatkozunk az utolsó η_t hivatkozás óta. LEWIS és SHEDLER [23] statisztikailag bebizonyították, hogy ezen modellek egyike sem teljesen megfelelő. Ebben a részben a független hivatkozási string modellt használjuk.

Feltesszük, hogy a program n lapból áll és csak $m(m<n)$ lehet ebből az első szinten. A független referencia modell a program viselkedésére a legegyszerűbb, melyet egyszerűsége és az elemzés kényelme miatt leggyakrabban használnak. Az i lap hivatkozásának a valószínűsége t időben p_i, minden t-re ($i=1, 2, \ldots, n$). Jelölje η_1, $\eta_2, \ldots, \eta_t, \ldots$ a hivatkozási stringet; a feltétel szerint ez egy független, azonos eloszlású valószínűségi változó sorozat:

$$P\{\eta_t = i\} = p_i, \quad i = 1, 2, \ldots, n.$$

Bármely laphelyettesítési algoritmusnál a teljesítmény mutatóul hosszantartó futás után a laphiba gyakoriság várható értékét használják sok esetben (lásd pl. AVEN, BOGUSLAVSKY, KOGAN [7], GELENBE [4]).

Korábbi cikkeinkben (ARATÓ [2], [3], BENCZÚR, KRÁMLI, PERGEL [8]) a teljesítmény mutatójául a lapolási hibák várható számát használtuk. Amikor a p_i eloszlás ismert az optimális laphelyettesítési algoritmus a nagy valószínűség értékekhez tartozó lapokat tartja bent. Az erről a problémáról szóló cikkek többségében (lásd pl. AHO ET AL. [1], FRANASZEK, WAGNER [13], AVEN ET AL. [7]) javasolt algoritmusok csak asszimptotikusan lesznek optimálisak, amikor a valószínűségi eloszlások nem ismertek és azokat a program végrehajtása során kell becsülni. Illetve a bizonyítások nem vihetők át triviálisan az ismeretlen paraméter esetére.

Ismeretlen p_i eloszlások esetén a lapolási hibák várható száma minimalizálásának feladatát először ARATÓ fogalmazta meg [2]. A *Bayes módszert* és az ún. „kétkarú bandita problémájának" a megoldását (lásd DE GROOT [12]) használta. A BAYES-I MÓDSZER esetében később bebizonyította [3], [4], hogy két lapból álló egyszerű esetben, a „legritkábban használt lap" kerül ki (az ún. *LFU*) stratégia optimális minden véges időintervallumra, ahol a hivatkozási string független azonos eloszlású valószínűségi változók sorozata, ismeretlen eloszlással. BENCZÚR, KRÁMLI, PERGEL [8] a veszteségfüggvény két szélsőséges esetében egy hasonló tételt tetszőleges n-re és m-re bizonyították. Ők szintén a *Bayes-féle módszert* használták, és általánosították a „két karú bandita problémájának" megoldását.

Most feltesszük, hogy a p_i valószínűségek ismertek, és az egyszerűség kedvéért $p_1 \geqq \ldots \geqq p_n \geqq 0$, $\sum_{i=1}^{n} p_i = 1$. Jelentse d_t a t idő után a központi memóriában jelenlevő lapok halmazát (d_t $n-m$ elemből áll). Feltesszük, hogy a d_t döntés csak a d_0 kezdeti döntéstől, és az $\{\eta_1, \ldots, \eta_{t'}\}$ $t' \geqq t$ megfigyelt hivatkozási stringektől függ. Jelöljük $D_{t,N}$-nel minden lehetséges szekvenciális $\{d_t, \ldots, d_{N-1}\}$ *Markov döntési eljárásnak* a halmazát, a véges $[t, N)$ intervallumon. Az első modellünkben, *A)* eset (lásd ARATÓ [2]), a memória átrendezhető minden $\eta_t \in \bar{d}_t$ hivatkozás után különösebb veszteség nélkül, de egy lapolási hiba, azaz $\eta_t \in d_t$, a költséget 1 egységgel növeli. Ez azt jelenti, hogy egy lap a második szintből átadódik az $m+1$ helyre és a lap tartalmának felhasználása után az elsőről egy lapot a második szintre kell helyezni.

A veszteség függvény a következő:

$$X_t^{d_{t-1}} = \begin{cases} 1, & \text{ha } \eta_t \in d_t \\ 0, & \text{egyébként.} \end{cases} \tag{2.1}$$

A másik esetet, *B)* eset (lásd ARATÓ [2]), a következő formában vizsgáltuk: egy lapcsere növeli a költséget a t pillanatban 1 egységgel, és η_t-nek a központi memóriában kell lennie. Ez azt jelenti, hogy az igényelt lap a második szintből bekerül egy másik helyére az első szinten, azaz két lapot ki kell cserélni.

A veszteség függvény ekkor a következő:

$$X^{d_t, d_{t-1}} = |d_t \setminus d_{t-1}|, \tag{2.2}$$

ahol $|\cdot|$ egy véges halmaz elemeinek számát jelöli. Megjegyezzük, hogy ha $\eta_t \in d_{t-1}$, akkor $X_t^{d_t, d_{t-1}} \geqq 1$. A célunk az, hogy megtaláljuk a *Markov típusú szekvenciális döntési eljárások*, $d_0, \ldots, d_{N-1}$ halmazát, amely minimalizálja a rizikó függvényt

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

(illetve

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

az *A)* illetve a *B)* esetnek megfelelően.

Jelöljük a feltételes várható értéket $\eta_1, \ldots, \eta_t$ egy adott $y_1, \ldots, y_t$ realizációja mellett $\mathbf{E}\{y_1, \ldots, y_t\}$-vel, és definiáljuk a rizikó-függvény családot a következő módon:

$$v(y_1, \ldots, y_t, N-t) = \min_{\{d_t, \ldots, d_{N-1}\} \in D_{t,N}} \mathbf{E}\{y_1, \ldots, y_t\} \sum_{\tau=t+1}^{N} X_\tau^{d_{\tau-1}} \tag{2.3}$$

és

$$v(y_1, \ldots, y_t, d_t, N-t) = \min_{\{d_{t+1}, \ldots, d_{N-1}\} \in D_{t+1,N}} \mathbf{E}\{y_1, \ldots, y_t\} \sum_{\tau=t+1}^{N} X_\tau^{d_\tau, d_{\tau-1}} \tag{2.4}$$

az *A)* illetve a *B)* esetben. A (2.3) és (2.4) függvénycsaládok kielégítik a *Bellmann egyenleteket* (lásd pl. [7]).

$$v(y_1, \ldots, y_{t-1}, N-t+1) = \min_{d_{t-1}} \mathbf{E}\{y_1, \ldots, y_{t-1}\}[X_t^{d_{t-1}} + \\ + v\{y_1, \ldots, y_{t-1}, \eta_t, N-t)], \tag{2.5}$$

illetve

$$v(y_1, \ldots, y_{t-1}, d_{t-1}, N-t+1) = \\ = \min_{d_t} \mathbf{E}\{y_2, \ldots, y_{t-1}\}[X_t^{d_t, d_{t-1}} + v(y_1, \ldots, y_{t-1}, \eta_t, N-t)]. \tag{2.6}$$

Ha rekurzive megoldjuk a (2.5) és (2.6) egyenletrendszert, megkapjuk az optimális stratégiákat. Az *A)* esetben $v(y_1, \ldots, y_t, N-t)$ nem függ d_{t-1}-től, ezért elég minimalizálni minden t-re az

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

feltételes várható értéket. Ugyanez az állítás, csekély módosítással igaz a *B)* esetben. Így

$$\min_{d_{t-1}} \mathrm{E}\{y_1, \ldots, y_{t-1}\}[X_t^{d_t-1}] = p_n + p_{n-1} + \ldots + p_{m+1} \tag{2.7}$$

és

$$\min_{d_t} \mathrm{E}\{y_1, \ldots, y_{t-1}\}[X_t^{d_t, d_t-1}] = (p_n + p_{n-1} + \ldots + p_{m+2} + p_l), \tag{2.8}$$

ahol

$$1 \leqq l \leqq m+1.$$

Így bebizonyítottuk a következő tételt, amely A. LEW [16] egy tételének az általánosítása.

2.1. TÉTEL. Feltéve, hogy az $\{\eta_t\}$, hivatkozási string független, stacionáris valószínűségi sorozat, a legkisebb hivatkozási valószínűséggel rendelkező lap második szintre helyezése az optimális helyettesítési politika.

Ebben a részben rövid összefoglalóját is adjuk korábbi eredményeinknek és modelljeinknek. Az $\eta_1, \eta_2, \ldots, \eta_t$ hivatkozási string valószínűségi szempontból független, azonos eloszlású valószínűségi változók sorozata; az η_t valószínűségi változók közös, $P_{i,w} = P_w(\eta_t = i)$, valószínűség eloszlása függ a w paramétertől, melynek értéke ismeretlen. A w-től való függés a következőképpen adott: a w paraméter értékkészlete az $\{1, \ldots, n\}$ természetes számok összes permutációjának W halmaza; $w(i)$ az $\{1, \ldots, n\}$ halmaz w-vel realizált egyértelmű leképezését jelöli. Feltesszük, hogy adott valószínűségek egy fix, csökkenő sorozata: $p_1 > p_2 > \ldots > p_n \geqq 0$ $((p_1 + \ldots + p_n) = 1)$ és

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

A döntéselméletben használt *Bayes módszer* alapján feltesszük, hogy w maga is egy valószínűségi változó. Mivel előzetes információnk nincs a $P_w\{\eta_t = i\}$ eloszlásról, feltesszük, hogy a w paraméter *apriori eloszlása* egyenletes.

Az *LFU stratégia* (a legkevésbé használt lap kerül ki) optimalitása a következő lemma következménye (lásd BENCZÚR, KRÁMLI, PERGEL [13]).

2.1. LEMMA. Ha a w paraméter *apriori eloszlása* egyenletes, és az *i.* lap f_i gyakorisága az $\{y_1, \ldots, y_{t-1}\}$ stringben kisebb, mint a *j.* lap f_j gyakorisága, akkor

$$P\{\eta_t = i | y_1, \ldots, y_{t-1}\} < P\{\eta_t = j | y_1, \ldots, y_{t-1}\}, \tag{2.7}$$

azaz a lapok aposteriori valószínűségeinek sorrendje az $\{y_1, \ldots, y_{t-1}\}$ string megfigyelése után ugyanaz, mint a gyakoriságuk sorrendje ebben a stringben.

Bizonyítás. Legyen i és j két olyan lap, hogy $f_i < f_j$. Ha w_1 és w_2 két permutáció a

$$w_1(i) = w_2(j), \tag{i}$$

$$w_2(i) = w_1(j), \tag{ii}$$

$$w_1(i) < w_1(j) \Leftrightarrow w_2(j) < w_2(i), \tag{iii}$$

$$w_1(k) = w_2(k), \quad k \neq i, j \tag{iv}$$

tulajdonságokkal, akkor a *Bayes tételt* használva:

$$P\{w_1|y_1,\dots,y_t\}=\frac{\prod_{k=1}^{n}P_{w_1(k)}^{f_i}}{\sum_{w\in W}\prod_{i=1}^{n}P_{w(i)}^{f_i}}$$

és

$$P\{w_1|y_1,\dots,y_t\}<P\{w_2|y_1,\dots,y_t\}.$$

Összegezve ezeket a valószínűségeket, megkapjuk a kívánt eredményt.

Az 1.1. lemmából és a w paraméter *apriori eloszlásának* egyenletességéből a következő állítást kapjuk.

2.2. TÉTEL. Az *LFU stratégia* minimalizálja az $\mathrm{E}\left[\sum_{t=1}^{N}X_t^{d_{t-1}}\right]$ várható veszteséget az *A)* esetben, ahol d_0 tetszőleges, és a w valószínűségi változó kezdeti ζ eloszlása egyenletes.

A *B)* esetben óvatosabban kell okoskodnunk. Először bebizonyítjuk, hogy az optimális stratégiák a lapoláshelyettesítési *(demand-page)* algoritmusok között vannak, azaz azon algoritmusok között, melyek kielégítik a

$$d_t=d_{t-1},\quad \text{ha}\quad \eta_t\notin d_{t-1},$$

$$d_{t-1}\setminus d_t=\{\eta_t\},\quad \text{ha}\quad \eta_t\in d_{t-1}$$

feltételeket.

Az *LFU stratégia* optimalitása a *B)* esetben a 2.3. tételből következik.

2.3. TÉTEL. Ha d_t és d_t' két különböző döntés, melyre

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

és a stringben az i-edik lap f_i gyakorisága kisebb, mint a j-edik lap f_j gyakorisága, akkor

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

A bizonyítás, $Q=N-t$-re indukcióval történhet. A 2.3. tétel állítása $Q=1$-re nyilvánvaló következménye az *A)* esetben használt megfigyeléseknek (1.1. lemma). Az indukciós lépés bizonyítása, a feltételes rizikó függvény

$$V(y_1,\dots,y_t,d_t,N-t)$$

és

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

összehasonlítása $t<N-1$ esetén nem olyan egyszerű, mint az *A)* esetben. Itt lényegesen kihasználjuk azt a tényt, hogy

$$v(y_1,y_2,\dots,y_t,d_t,N-t)$$

az $\{y_1,y_2,\dots,y_{t'}\}$ stringekben csak a lapok gyakoriságától és d_t-től függ. Részletesebb bizonyítás található [8]-ban.

Megjegyezzük, hogy *LFU stratégia* optimalitása a *Bayes feltétel* nélkül szintén igaz marad (lásd ARATÓ, BENCZÚR, KRÁMLI [5]).

3. Egy modell tranzakciókra adatbázis rendszerben

Egy számítógép által kezelt nagy adatbázis rendszerhez több felhasználó is hozzáférhet, pl. távoli terminálokról. Egy ilyen rendszer viselkedésének leírása és a feldolgozás jellemzése tipikus és központi probléma egy adatbázis rendszer értékelésekor. Egy ilyen rendszert a lehető leggazdaságosabban szeretnénk megszervezni, oly módon, hogy egy kérdésre megfelelően rövid idő alatt kapjunk választ adott felhasználói szám mellett. A feladatnál figyelembe vesszük a várakozási költséget és a számítógép rendszer erőforrásainak kihasználását is. Az adatbázis rendszer ily módon komplex és dinamikus, ezért az értékelés végrehajtása előtt, amely magában foglalhatja a beállítást és az előrejelzést, mérésekre van szükség az adatgyűjtéssel, elemzéssel, modellezéssel és interpretálással kapcsolatban. Ilyen méréseket és statisztikai elemzéseket készített LEWIS és SHEDLER [18], hogy megkapják az *Információ Kezelő Rendszer (IMS)* viselkedésének egy matematikai leírását. Az *IMS*-ben a felhasználók távoli terminálokból léphetnek be az adatbázisba, a belépő közléseket *tranzakcióknak* nevezzük. Egy bizonyos tranzakció egy alkalmazási programot használ, amely feldolgozza a tranzakciót és belép az adatbázisba. Az *IMS* adatkezelő képességét *Adat Nyelv/I*-nek *(DL/I)* nevezzük. Egy alkalmazási program végrehajtása így hívások egy sorozata, az *IMS DL/I* komponenséhez.

LEWIS és SHEDLER [18] cikkükben elemezték a tranzakció kezdés folyamatát. A modellük a következő volt: Az események számát (tranzakció kezdések) $[0, t]$-ben, N_t-vel jelölték, ahol az N_t várható értéke $M(t)=\mathbf{E}(N_t)$.

Az N_t folyamatra feltették, hogy nem-homogén *Poisson-folyamat* (lásd pl. B. GNEDENKO, I. KOVALENKO [11] vagy D. COX, P. LEWIS [10]), mely azt jelenti, hogy

$$(3.1)\qquad P\{N(t)-N(s)=k\} = \frac{e^{-[\Lambda(t)-\Lambda(s)]}[\Lambda(t)-\Lambda(s)]^k}{k!}.$$

$$\Lambda(t)-\Lambda(s) = \int_s^t \lambda(u)\,du$$

($\lambda(u)\geqq 0$, a sűrűségfüggvény) és

$$P\{N(t)-N(s)=0\} = 1-\lambda(t)(t-s)+o(t-s),$$

$$P\{N(t)-N(s)=1\} = \lambda(t)(t-s)+o(t-s),$$

$$P\{N(t)-N(s)=2\} = o(t-s).$$

Véges sok, diszjunkt intervallumon az események számai független valószínűség változók. Könnyen látható, hogy

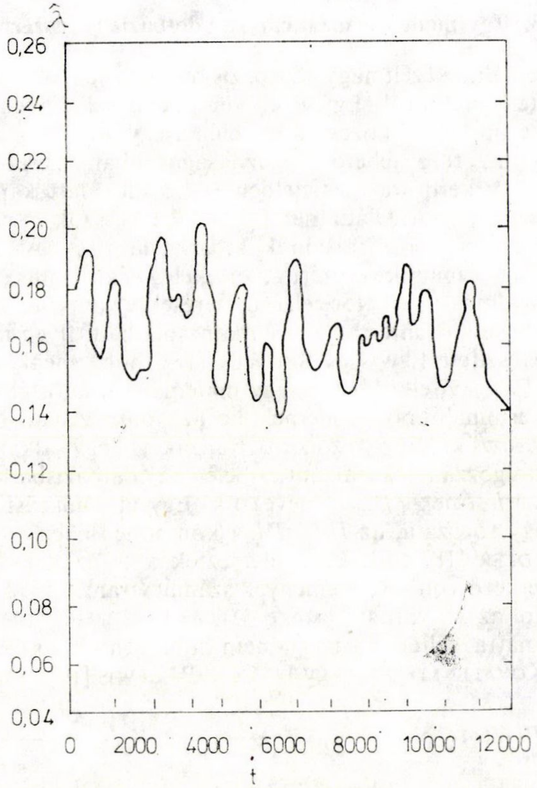
$$\Lambda(t) = \mathbf{E}\{N_t\}.$$

LEWIS és SHEDLER azt találták, hogy az adatok szembetűnő tulajdonsága a beérkezési sűrűség függvény oszcilláló természete volt. A spektrál analízis nem-stacionárius adatokra nem volt alkalmazható.

A $\lambda(t)$ beérkezési sűrűség függvényt egy exponenciális polinomként becsülték, a következőképpen:

$$(3.2)\qquad \lambda(t) = \exp\left\{\sum_{m=1}^{r} \alpha_m t^m\right\}.$$

A $\lambda(t)$ függvény tipikus becsléseit az olvasó megtalálhatja az 1., 2. ábrákon.



1. ábra

A determinisztikus beérkezési sűrűség függvény közelítése csak első lépésnek tekinthető. Ha több oszcillációs hatásra számítunk egy nap alatt, hatékonyabbnak tűnik egy duplán sztochasztikus modell használata, melyet ARATÓ, KNUTH, TŐKE [6] (lásd még KNUTH [15]) multiprogramozású rendszerek leírására javasoltak. SNYDER [20] hasonló modellt használt duplán sztochasztikus *Poisson folyamatok* szűrésére és értékelésére. Ezt a folyamatot először COX [9] vezette be és COX és LEWIS [10] írták le.

Az általános matematikai modellben a sebesség függvényt (vagy a paramétereket) a $\lambda(t)=h(t, X(t))$ formában tekintettük, ahol $X(t)$ sztochasztikusan változik időben, pl. mint egy *Markov diffúziós folyamat*. A tranzakció belépéseket számláló folyamat intenzitása sztochasztikusan változik két ok miatt, egyik a napi oszcilláció, a másik a programok (vagy jobok) változása. Feltesszük, hogy az N_t folyamat, $\{\lambda(t), t>t_0\}$ sztochasztikus sűrűségfüggvénnyel, a következő tulajdonságú:

$$N_0 \equiv 0 \quad \text{(1 valószínűséggel)},$$

$$P\{N_t - N_s = k \mid \lambda(\sigma),\ s<\sigma<t\} = \frac{\left[\int_s^t \lambda(\sigma)\,d\sigma\right]^k}{k!}\, e^{-\int_s^t \lambda(\sigma)\,d\sigma} \tag{3.3}$$

és

$$P\{N_{t+\Delta t}-N_t = i|\lambda(\sigma),\ t<\sigma<t+\Delta t\} =$$
$$= \big(1-\lambda(t)\Delta t\big)\cdot\delta_{0i}+\lambda(t)\Delta t\delta_{1i}+o(\Delta t),$$

$$m_t = \int_0^t \lambda(\sigma)\,d\sigma = \mathbf{E}\{N_t|\lambda(\sigma),\ 0<\sigma<t\}.$$

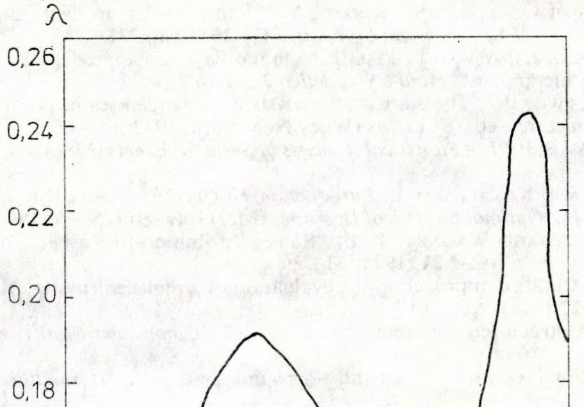
Feltesszük továbbá, hogy a $\xi(t)$ folyamat a következő alakú

(3.4) $$d\xi_t^{(l-1)}+[a_0\xi_t+a_1\xi_t'+\ldots+a_1\xi_t^{(l-1)}]\,dt = dw(t),$$

ahol $l\geqq 1$, $w(t)$ *Brown mozgás folyamat (Wiener-folyamat)*. Legyen

$$\lambda(t) = |\xi_t|$$

és tekintsük a $\lambda(t)$ becslésének problémáját az N_u számláló folyamat megfigyeléséből a $(0, t)$ intervallumon. A becslési probléma megoldása függ a ξ_t *aposteriori eloszlásától* (vagy sűrűségétől). Jelölje $L[\cdot]$ a ξ_t *diffúziós folyamatra* vonatkozó *Kolmogorov féle első differenciál operátort*, továbbá $\hat{\lambda}(t)=\mathbf{E}[\lambda_t(\xi_t)|N_t]$ a $\lambda_t(\xi_t)$ sűrűség négyzetes középben legjobb becslését a determinisztikus esetben, adott N_t esetén.



2. ábra

ξ_t aposteriori $p_t(X_t|N_t)$ sűrűségére a következő differenciálegyenletet kapjuk (lásd SNYDER [20])

$$(3.5) \quad dp_t(\xi_t|N_t) = L[p_t(\xi_t|N_t)]dt + p_t(\xi_t|N_t)\cdot\{\lambda_t(\xi_t)-\hat{\lambda}_t\}\cdot\lambda_t^{-1}\{dN_t-\hat{\lambda}_t\,dt\}.$$

Az egyenlet megoldását csakis numerikus számolással kaphatjuk meg.

IRODALOM

[1] AHO, A. V., DENNING, P. J. AND ULLMANN, J. D., "Principles of optimal page replacement", *Journal A. C. M.* **18** (1971) 80—93.

[2] ARATÓ, M., „Számítógépek hierarchikus laptárolási eljárásainak optimalizálásáról", *MTA SZTAKI Közlemények* **16** (1976) 7—23.

[3] ARATÓ, M., "Statistical sequential methods in performance evaluation of computer system", *in 2nd International workshop on modelling and performance evalaution of computer systems* (Stresa-Italy, 1976) 1—10.

[4] ARATÓ, M., "A note on optimal performance of page storage", *Acta Cybernetica* **3** (1976) 25—30.

[5] ARATÓ, M., BENCZÚR, A. AND KRÁMLI, A., "On the solution of optimal performance of page storage hierarchies with independent reference string", Banach Center Publications, Mathematical Statistics, **6** (1977), megjelenés alatt.

[6] ARATÓ, M., KNUTH, E. AND TŐKE, P., "On stochastic control of a multiprogrammed computer based on a probability model", *in Preprints of the Stochastic Control Symposium* (IFAC, Budapest, 1974) 305—313.

[7] AVEN, O., BOGUSLAVSKY, L. AND KOGAN, Y., "Some results on distribution-free analysis of paging algorithms", *IEEE Trans. on Computers* **C—25** (1976) 737—745.

[8] BENCZÚR, A., KRÁMLI, A. AND PERGEL, J., "On the Bayesian approach to optimal performance of page storage hierarchies", *Acta Cybernetica* **3** (1976).

[9] COX, D. AND LEWIS, P., "The statistical analysis of dependencies in point processes", *in Stochastic Point Processes*, ed.: P. LEWIS (Wiley New York, 1972) 55—66.

[10] COX, D. AND LEWIS, P., *The Statistical Analysis of Series of Events* (Methuen, London and Wiley, New York, 1966).

[11] GNEDENKO, B. AND KOVALENKO, I., *Introduction To Queuing Theory* (Moscow, 1969).

[12] DE GROOT, M. H., *Optimal Statistical Decisions* (Mc. Graw-Hill, N. Y., 1970).

[13] FRANASZEK, P. A. AND WAGNER, T. J., "Some distribution-free aspects of paging algorithm performance", *Journal ACM* **21** (1974) 31—39.

[14] GELENBE, E., "A unified approach to the evaluation of a replacement algorithms", *IEEE Trans. Computers* **C—22** (1973) 611—617.

[15] KNUTH, E., "A structured computer system model", *Comp. and Maths. With Appl.* **1** (1975) 327—336.

[16] LEW, A., "Optimal control of demand — paging systems", *Information Sciences* **10** (1976) 319—330.

[17] LEWIS, P. AND SHEDLER, G., "Empirically derived micromodels for sequences of page exceptions". *IBM J. Res. Developm.* 1973) 86—100.

[18 LEWIS, P. AND SHEDLER, G., "Statistical analysis of transaction processing in a data base system", Sept. R. J. 1629. *IBM Research*, (1975).

[19] SEGALL, A., "Dynamic file assignment in a computer network", *IEEE Trans. on Automatic Control* **AC—21** (1976) 161—173.

[20] SNYDER, D., "Filtering and detection for doubly stochastic Poisson processes", *IEEE Trans. on Inform. Theory* **18** [1975) 91—102.

(Beérkezett: 1977. szeptember 16.)

ARATÓ MÁTYÁS
SZÁMÍTÓGÉPALKALMAZÁSI KUTATÓ INTÉZET
1536 BUDAPEST, POSTAFIÓK 227.

STATISTICAL SEQUENTIAL DECISION METHODS IN CASE OF INDEPENDENT STRINGS

M. ARATÓ

The purpose of this paper is to find optimal policies in exact forms for dynamical allocation of files, for page replacement and for estimation of rate parameter of transactions.

In this work relatively simple models are considered to present the main method of solving such problems. We are using the well known method of statistical sequential decision theory. In addition to the application-oriented contribution in analysis of such systems our results contain also theoretical contribution in the theory of sequential methods (or optimal control).

Section 1. presents the basic model used in file assignement in computer networks with independent request strings. The problem is solved in the papers of A. SEGALL, but he did not get simple solutions because his model is more complicated. Section 2. contains our earlier results in paging when the reference string is an independent sequence.

In Section 3. we are giving a new model for transactions in a data base system, the statistical description and analysis of which was done by LEWIS and SHEDLER.