線形法におけるバケット方式を用いたときのアクセス回数について†

中村良三†† 松山公一†††

分散記憶法を2次記憶系での探索に利用するとき，ファイルへのアクセス方法としては，複数個の見出しをひとまとめにして取り扱うバケット方式がある．この方式で，あふれの処理に線形法を用いたとき，アクセス回数に関して Knuth が提案した評価式は，見出しの探索頻度が一様であるとの仮定の下で，アクセス回数の近似値を評価するものである．本論文では，各見出しの探索頻度を考慮する立場から，アクセス回数を評価する表現式を導き出し，探索頻度に具体的な確率分布を与えたときのアクセス回数を評価する．とくに，探索頻度を一様と仮定したとき，従来の評価式では，なぜアクセス回数の厳密な評価でなく近似値を評価するのか，その導出過程を示し検討する．また，あふれの処理として，分離連鎖法と線形法を用いたときのアクセス回数を評価した結果，バケットサイズが大きくかつ表占有率が大きい場合には，線形法が分離連鎖法よりも探索効率がよいことを示す．

1. まえがき

分散記憶法を2次記憶系での探索に利用するとき，ファイルへのアクセス方法としては，複数個の見出しをひとまとめにして取り扱うバケット方式がある．このあふれの処理では，連続した探索路ができるだけ同一ページ内にあり，新たなページを読み込む必要が少ないことが望ましい．したがって，衝突処理としては，分離連鎖法や線形法が有効である．分離連鎖法を用いた場合のアクセス回数の評価については，すでに文献1),4)で論議されている．線形法の場合には，Knuth によって，Schay-Spruth のモデル[2)] をバケット方式に拡張し，アクセス回数を評価する方法が，文献1)で考察されている．

しかし，この Schay-Spruth のモデルは，線形法における探索路長の近似値を評価するものである．したがって，本論文では，衝突処理として線形法を用いたとき，各見出しの探索頻度を考慮した探索路長を，厳密な解析の下で評価した文献3)のモデルをバケット方式に拡張し，アクセス回数を評価する表現式を導き出す．次に，この評価式を用い，各見出しの探索頻度に具体的な確率分布を与えたときのアクセス回数を評価する．とくに，探索頻度が一様である場合，Knuth の表現式によって評価されたアクセス回数と比較検討する．また，線形法と分離連鎖法でのアクセス回数を評価し比較する．

なお，評価式の導出過程で，分散表の大きさを M，バケットサイズを b，見出しの数を N，表占有率を $\alpha(=N/Mb)$ とする．また，各見出しが探索される確率を，登録順序に従い $p_i(i=1,2,\cdots,N)$ とし，探索が成功するときの平均アクセス回数を S_N，分散を V_N，不成功のときのそれらをそれぞれ $\bar{S}_N, \bar{V}_N$ とする．

2. 提案する評価式

線形法でのあふれの処理は分離連鎖法のようにポインタで連結しないで，隣接する空きのあるバケットを探し見出しを登録する．すなわち，見出しの値 K は，分散関数 $h(K)$ によって，0 から $M-1$ の分散番地のいずれかの番地に変換され，その分散番地のバケットにまだ空きがあれば，そのバケットに見出しを登録する．しかし，そのバケットが満状態であれば，次の登録候補番地として $h(K)-1$ 番地のバケットを調べ，そのバケットがまた満状態であれば，さらに，$h(K)-2$ 番地のバケットと順次に，空きがあるバケットを探し，そのバケットに見出しを登録する．このとき，表の始端すなわち分散番地が0のバケットの次は，表の終端の $M-1$ 番地のバケットへと巡回して登録場所を探すものとする．

一方，このようにして登録された見出しの探索には，登録されたと同じ探索路を辿りながら，バケット単位にアクセスする．

ここで，バケット方式でのアクセス回数を考察するに先だって，分散番地0のバケットから分散番地 $M-1$ のバケットまで，順次バケット内の見出しをふくむレコードに通し番号 $(0,1,\cdots,Mb-1)$ を付け，バケットサイズが1で分散表の大きさが Mb であるよう

† On the Number of Accesses for the Bucket Storage Technique in Linear Probing by RYOZO NAKAMURA (Faculty of Engineering, Kumamoto University) and KIMIKAZU MATSUYAMA (Kumamoto University).

†† 熊本大学電子計算機室
††† 熊本大学

な場合を想定すれば，そのときのアクセス回数，すなわち探索路長は，文献3) で厳密な解析の下で評価されている．

したがって，文献3) に従い，大きさ Mb の分散表に，n 個の見出しが登録されている状態で，次の関数および確率を導入する．

$f(Mb,n)$：表の 0 番地が空きであるような分散番地列の数．

$g(Mb,n,h)$：表の 0 番地が空きで，1 番地から h 番地まで見出しで占有され，$h+1$ 番地が空きであるような分散番地列の数．

P_{nj}：$n+1$ 番目の見出しを登録するとき，$j+1$ 回の探索を要する確率．

このとき，これらの関数および確率は次のように表現される．

$$f(Mb,n)=\left(1-\frac{n}{Mb}\right)(Mb)^n \qquad (1)$$

$$g(Mb,n,h)=\binom{n}{h}(h+1)^{h-1}(Mb-h-1)^{n-h-1} \times (Mb-n-1) \qquad (2)$$

$$P_{nj}=(Mb)^{-n}\{g(Mb,n,j)+g(Mb,n,j+1)+\cdots+g(Mb,n,n)\} \qquad (3)$$

本題に戻り，大きさ Mb の分散表をバケットサイズ b によって，M 個のバケットに区分し直し，バケット方式におけるアクセス回数を考察する．この際，新たに次の記号を導入する．

Q_{nk}：$n+1$ 番目の見出しを登録するとき，$k+1$ 回のバケットの探索を要する確率．

すなわち，$n+1$ 番目の見出しを登録するとき，バケットをただ1回調べることによって登録される確率 Q_{n0} は，前述の表現を用いれば，探索回数が1からバケットの大きさ b までの場合に対応する．

したがって，

$$Q_{n0}=P_{n0}+P_{n1}+\cdots+P_{n,b-1}$$

となる．同様に，バケットを2回探索して登録される確率 Q_{n1} は，

$$Q_{n1}=P_{n,b}+P_{n,b+1}+\cdots+P_{n,2b-1}$$

となる．よって，一般的に Q_{nk} と P_{nj} の間には次の関係が成立する．

$$Q_{nk}=\sum_{j=kb}^{(k+1)b-1}P_{nj},\quad \left(k=0,1,\cdots,\left\lceil\frac{N}{b}\right\rceil^{*}-1\right) \qquad (4)$$

一方，N個の見出しが登録されている分散表の探索は，登録されたときと同じ探索路を辿るので，バケットを k 回アクセスすることによって見つかる見出しは，登録の際にも同じ k 回のバケットの探索を要して登録された見出しである．

したがって，見出しが k 回のアクセスで探索される確率は，

$$\rho_{(k-1)b+1}Q_{(k-1)b,k-1}+\rho_{(k-1)b+2}Q_{(k-1)b+1,k-1}+\cdots+\rho_N Q_{N-1,k-1},\quad \left(k=1,2,\cdots,\left\lceil\frac{N}{b}\right\rceil\right)$$

となる．

ここで，アクセス回数を確率変数に設定すれば，成功探索時のアクセス回数の平均，分散は次のように表現される．

$$S_N=\sum_{k=0}^{\lceil N/b\rceil-1}(k+1)\sum_{j=kb}^{N-1}\rho_{j+1}Q_{jk}$$
$$=\sum_{k=0}^{\lceil N/b\rceil-1}(k+1)\sum_{j=kb}^{N-1}\rho_{j+1}\sum_{i=kb}^{(k+1)b-1}P_{ji}$$
$$=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}\left(\left\lfloor\frac{j}{b}\right\rfloor^{*}+1\right)P_{kj} \qquad (5)$$

$$V_N=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}\left(\left\lfloor\frac{j}{b}\right\rfloor+1\right)^2 P_{kj}-S_N^2 \qquad (6)$$

不成功探索時のアクセス回数の平均，分散は $N+1$ 番目の見出しを登録するときに要するバケットの探索回数の平均，分散と同値であるから，それらは次のように表現することができる．

$$\bar{S}_N=\sum_{k=0}^{N}\left(\left\lfloor\frac{k}{b}\right\rfloor+1\right)P_{Nk} \qquad (7)$$

$$\bar{V}_N=\sum_{k=0}^{N}\left(\left\lfloor\frac{k}{b}\right\rfloor+1\right)^2 P_{Nk}-\bar{S}_N^2 \qquad (8)$$

3. 比較検討

各見出しの探索頻度に具体的な確率分布を与えたときのアクセス回数について論議する．とくに，探索頻度が一様な場合には，提案する評価式と Schay-Spruth のモデルを拡張して導出された Knuth の評価式と比較し，その相違点を考察する．

次に，分離連鎖法と線形法におけるアクセス回数を評価し比較する．

3.1 探索頻度が一様なとき

a）提案する評価式の場合

探索頻度が一様という仮定から，$\rho_i=1/N,(i=1,\cdots,N)$ となり，(5), (6)からアクセス回数の平均，分散は次のようになる．

---

* $\lceil x\rceil$ はxよりも大きい最小の整数を表す．

* $\lfloor x\rfloor$ はxよりも大きくない最大の整数を表す．

$$S_N=\frac{1}{N}\sum_{k=0}^{N-1}\sum_{j=0}^{k}\left(\left\lfloor\frac{j}{b}\right\rfloor+1\right)P_{kj}$$
$$=1+\frac{1}{N}\left(\sum_{k=b}^{N-1}\sum_{j=b}^{k}P_{kj}+\sum_{k=2b}^{N-1}\sum_{j=2b}^{k}P_{kj}+\cdots+\sum_{k=\lambda b}^{N-1}\sum_{j=\lambda b}^{k}{}^{*}P_{kj}\right)$$
$$=1+\frac{1}{N}(f_b+f_{2b}+\cdots+f_{\lambda b}) \quad (9)$$

$$V_N=1+\frac{1}{N}\left(3\sum_{k=b}^{N-1}\sum_{j=b}^{k}P_{kj}+5\sum_{k=2b}^{N-1}\sum_{j=2b}^{k}P_{kj}+\cdots+(2\lambda+1)\sum_{k=\lambda b}^{N-1}\sum_{j=\lambda b}^{k}P_{kj}\right)-S_N{}^2$$
$$=\frac{1}{N}(f_b+3f_{2b}+5f_{3b}+\cdots+(2\lambda-1)f_{\lambda b})-\frac{1}{N^2}(f_b+f_{2b}+\cdots+f_{\lambda b})^2 \quad (10)$$

ここで，$f_b, f_{2b}, \cdots$ 等は次のように定義される．見出しの探索において，f_b はバケットを1回以上アクセスする確率の和を表す．同様に，f_{2b} はバケットを2回以上アクセスする確率の和を表す．したがって，見出しの探索において，バケットを k 回以上アクセスする確率の和，すなわち f_{kb} は次のように表現できる．

$$f_{kb}=\sum_{i=kb}^{N-1}\sum_{j=kb}^{i}P_{ij}$$
$$=\sum_{i=kb}^{N-1}(Mb)^{-i}\sum_{j=kb}^{i}(j-kb+1)g(Mb,i,j)$$
$$=\sum_{i=kb}^{N-1}\sum_{j=kb}^{i}(j-kb+1)\binom{i}{j}\left(\frac{j+1}{Mb}\right)^{j-1}\times\left(1-\frac{j+1}{Mb}\right)^{i-j-1}\left(1-\frac{i+1}{Mb}\right)\Big/Mb$$
$$(k=1,2,\cdots,\lambda) \quad (11)$$

b) Schay-Spruth モデルの拡張の場合

N 個の見出しが分散関数によって，分散番地 j のバケットに分散される見出しの個数を α_j，j 番地のバケットからあふれて $j-1$ 番地のバケットに移される見出しの数を β_j とする．このとき，α_j, β_j には次のような関係が成立する．

$$\left.\begin{array}{l}\alpha_0+\alpha_1+\cdots+\alpha_{M-1}=N,\\ \beta_j=\alpha_j+\beta_{(j+1)\bmod M}-b,^{**}\\ \beta_j\le N-b,\ (j=0,1,\cdots,M-1)\end{array}\right\} \quad (12)$$

また，α_j の値が k になる確率を p_k，β_j の値が k になる確率を r_k とする．ここでは，N 個の見出しが大きさ M の分散表に一様に分散されると考えているので，p_k は

$$p_k=\binom{N}{k}\left(\frac{1}{M}\right)^k\left(1-\frac{1}{M}\right)^{N-k} \quad (13)$$

となる．さらに，数列 p_k および r_k によって構成される母関数を，それぞれ $B(z), C(z)$ とすれば，それらは次のようになる．

$$B(z)=p_0+p_1z+p_2z^2+\cdots \quad (14)$$
$$C(z)=r_0+r_1z+r_2z^2+\cdots \quad (15)$$

ところで，成功時の平均アクセス回数 S_N は

$$S_N=1+\frac{M}{N}\sum_{k=0}^{N-b}kr_k$$
$$=1+\frac{M}{N}C'(1) \quad (16)$$

と表せる．

したがって，母関数 $C(z)$ の導関数 $C'(1)$ の表現が必要となる．

はじめに，母関数 $B(z), C(z)$ の積は

$$B(z)C(z)=p_0r_0+(p_0r_1+p_1r_0)z+(p_0r_2+p_1r_1+p_2r_0)z^2+\cdots \quad (17)$$

となる．

すなわち，たたみこみの定義から，$B(z)C(z)$ は

$$s_n=\sum_{0\le k\le n}p_kr_{n-k} \quad (18)$$

で定められる数列，$s_0, s_1, \cdots$ の母関数となる．このとき，Schay-Spruth のモデルに準じ

$$\left.\begin{array}{l}r_0=s_0+s_1+\cdots+s_b\\ r_1=s_{b+1}\\ r_2=s_{b+2}\\ \cdots\cdots\end{array}\right\} \quad (19)$$

と仮定すると，$B(z)C(z)$ は次のように表現される．

$$B(z)C(z)=Q(z)+z^bC(z) \quad (20)$$

ただし，

$$Q(z)=(1-z)\{p_0r_0+(p_0r_0+p_0r_1+p_1r_0)z+\cdots+(p_0r_0+\cdots+p_0r_{b-1})z^{b-1}\} \quad (21)$$

したがって，母関数 $C(z)$ は

$$C(z)=Q(z)/\{B(z)-z^b\} \quad (22)$$

となる．

ここで，$Q(z)=0$ の解を $1, q_1, q_2, \cdots, q_{b-1}$ とすれば，(22) は

$$C(z)=\frac{(z-1)(z-q_1)(z-q_2)\cdots(z-q_{b-1})}{B(z)-z^b} \quad (23)$$

となる．

次に，$B(1)=1$ であるから，新しい母関数 $D(z)$ を導入して，$B(z)$ を次のように表す．

---

* $\lambda\equiv\left\lfloor\frac{N-1}{b}\right\rfloor$ を表す．

** $\alpha_j+\beta_{(j+1)\bmod M}\le b$ のとき $\beta_j=0$.

$$B(z)=1+(z-1)D(z) \tag{24}$$

これを(23)に代入すると，$C(z)$ は

$$C(z)=\frac{(z-q_1)(z-q_2)\cdots(z-q_{b-1})}{D(z)-(z^{b-1}+z^{b-2}+\cdots+1)} \tag{25}$$

となる．したがって，$C'(1)$ は次のようになる．

$$C'(1)=\frac{(1-q_2)(1-q_3)\cdots(1-q_{b-1})+\cdots+(1-q_1)(1-q_2)\cdots(1-q_{b-2})}{D(1)-b} - \frac{(1-q_1)\cdots(1-q_{b-1})\{D'(1)-((b-1)+(b-2)+\cdots+1)\}}{\{D(1)-b\}^2} \tag{26}$$

一方，(25)から，

$$D(1)-b=(1-q_1)(1-q_2)\cdots(1-q_{b-1})$$

が成立するので，(26)は次のようになる．

$$C'(1)=\frac{1}{1-q_1}+\frac{1}{1-q_2}+\cdots+\frac{1}{1-q_{b-1}}-\frac{D'(1)-\frac{b(b-1)}{2}}{D(1)-b} \tag{27}$$

ここで，$D(1), D'(1)$ は(24)から，

$$D(1)=B'(1), D'(1)=B''(1)/2,$$

となるので，(27)は次のようになる．

$$C'(1)=\frac{1}{1-q_1}+\frac{1}{1-q_2}+\cdots+\frac{1}{1-q_{b-1}}-\frac{B''(1)-b(b-1)}{2\{B'(1)-b\}} \tag{28}$$

次に，$q_1, q_2, \cdots, q_{b-1}$ を求める．

初めに，$B(z)$ をポアソン分布の母関数

$$P(z)=e^{\alpha b(z-1)}$$

で近似すると，(22)は次のように表現できる．

$$Q(z)\fallingdotseq C(z)\{P(z)-z^b\} \tag{29}$$

ここで，Knuth は，$Q(z)=0$ を解くかわりに，(29)から，$P(z)-z^b=0$ を解き，種々の展開を行ったあと，$t_n(\alpha)$ なる関数を次のように，文献1)で導出している．

$$t_n(\alpha)=\frac{(-1)^{n-1}(n!\alpha)^{-1}\sum_{M>n}(-n\alpha)^M}{M(M-1)(M-n-1)!} \tag{30}$$

そして，アクセス回数の平均 S_N を次のように表現している．

$$S_N\fallingdotseq 1+t_b(\alpha)+t_{2b}(\alpha)+t_{3b}(\alpha)+\cdots \tag{31}$$

しかし，上述した導出過程では，次のような仮定がなされていると考えられる．

すなわち，(12)で示したように，α_i と β_i には従属な関係が存在するにもかかわらず，(19)の仮定では独立な確率変数として取り扱っている．

また，$\beta_i \leq N-b$ なる制約に伴い，確率 r_k は，$k>N-b$ のとき，$r_k=0$ となる制限が生ずるが，(19)では，それが加味されていない．

したがって，(31)はバケットサイズ b，表占有率 α を用いて簡潔に表現された評価式ではあるが，上述のような問題を含んでいるため，厳密な評価式とはなりえない．

c) 数 値 例

提案する評価式(9)から求めた平均成功アクセス回数と Knuth の評価式(31)を用いて評価された平均アクセス回数の数値例を **表 1***，**表 2** に示す．また，提案する評価式(10)によって評価された分散を **表 3*** に示す．

提案する評価式(9)と Knuth の評価式(31)によって評価した平均アクセス回数は，後者のほうの評価値が若干大きく，とくに表占有率が1の近傍では，その

表 1 提案する評価式による一様探索時の平均成功アクセス回数

Table 1 Average number of accesses in a successful search for uniform probing by the proposed formula.

Bucket size (b)	Load factor (α)									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	1.0425	1.1066	1.1875	1.2921	1.4323	1.6282	1.9176	2.3771	3.1797	3.9881
2	1.0035	1.0169	1.0419	1.0818	1.1434	1.2394	1.3961	1.6736	2.2369	2.7651
3	1.0004	1.0038	1.0132	1.0320	1.0656	1.1235	1.2260	1.4218	1.8597	2.3533
4	1.0001	1.0009	1.0048	1.0146	1.0347	1.0732	1.1463	1.2953	1.6536	2.0531
5	1.0000	1.0003	1.0019	1.0073	1.0200	1.0469	1.1019	1.2204	1.5236	1.9025
10	1.0000	1.0000	1.0000	1.0004	1.0023	1.0089	1.0257	1.0531	1.0752	1.0790
20	1.0000	1.0000	1.0000	1.0000	1.0001	1.0007	1.0028	1.0060	1.0084	1.0088

* 分散表の大きさ M を50とする．

表 2 Knuth の評価式を用いたときの平均成功アクセス回数
Table 2 Average number of accesses in a successful search by Knuth's formula.

Bucket size (b)	Load factor (α) 10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	1.0556	1.1250	1.2143	1.3333	1.5000	1.7500	2.167	3.000	5.500	10.5
2	1.0062	1.0242	1.0553	1.1033	1.1767	1.2930	1.494	1.903	3.147	5.6
3	1.0009	1.0066	1.0201	1.0450	1.0872	1.1584	1.286	1.554	2.378	4.0
4	1.0001	1.0021	1.0085	1.0227	1.0497	1.0984	1.190	1.386	2.000	3.2
5	1.0000	1.0007	1.0039	1.0124	1.0307	1.0661	1.136	1.289	1.777	2.7
10	1.0000	1.0000	1.0001	1.0011	1.0047	1.0154	1.042	1.110	1.345	1.8
20	1.0000	1.0000	1.0000	1.0000	1.0003	1.0020	1.010	1.036	1.144	1.4

表 3 提案する評価式による一様探索時の成功アクセス回数の分散
Table 3 Variance of accesses in a successful search for uniform probing by the proposed formula.

Bucket size (b)	Load factor (α) 10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	0.0459	0.1358	0.2874	0.5539	1.0474	2.0209	4.0997	8.9902	21.9287	39.8764
2	0.0036	0.0184	0.0511	0.1178	0.2566	0.5645	1.3216	3.4757	10.9753	21.4209
3	0.0004	0.0039	0.0146	0.0402	0.0993	0.2419	0.6246	1.8438	6.8881	16.0647
4	0.0000	0.0010	0.0051	0.0169	0.0472	0.1259	0.3522	1.1343	4.8062	11.6501
5	0.0000	0.0003	0.0020	0.0080	0.0253	0.0734	0.2202	0.7621	3.5725	9.7717
10	0.0000	0.0000	0.0000	0.0004	0.0025	0.0107	0.0360	0.0806	0.1169	0.1232
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007	0.0028	0.0060	0.0084	0.0087

表 4 見出しの探索頻度が登録順序に従い半減する場合の成功アクセス回数
Table 4 The number of accesses in a successful search for the reduction by half.

Bucket size	Load factor (α) 10%		30%		50%		70%		90%	
(b)	S_N	V_N	S_N	V_N	S_N	V_N	S_N	V_N	S_N	V_N
1	1.0174	0.0184	1.0213	0.0237	1.0213	0.0238	1.0213	0.0238	1.0213	0.0238
2	1.0003	0.0003	1.0003	0.0003	1.0003	0.0003	1.0003	0.0003	1.0003	0.0003
3	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
4	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
5	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
10	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
20	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000

差が顕著になる．

3.2 探索頻度に確率分布を与えたとき

a）探索頻度が登録順序に従い半減する場合

このとき，i 番目に登録された見出しが探索される確率 ρ_i は

$$\rho_i=\frac{1}{2^{i-1}(2-2^{1-N})},\quad (i=1,2,\cdots,N) \tag{32}$$

となる．

b）探索頻度が登録順序に従い調和減少する場合

このとき，ρ_i は次のようになる．

$$\rho_i=\frac{1}{iH_N},\quad (i=1,2,\cdots,N) \tag{33}$$

ただし，H_N は調和数で，$1+1/2+\cdots+1/N$ を表す．探索頻度を考慮したときの平均成功アクセス回数および分散は，(32)，(33)を(5)，(6)に，それぞれ代入することによって評価できる．

これらの数値例を**表 4**，**表 5** に示す．ただし，分散表の大きさを 50 とする．

3.3 線形法と分離連鎖法間の比較

2 次記憶を対象にした探索に，分散記憶法を適用す

表 5 見出しの探索頻度が登録順序に従い調和減少する場合の成功アクセス回数
Table 5 The number of accesses in a successful search for the harmonic reduction.

Bucket size (b)	Load factor (α) 10% S_N	10% V_N	30% S_N	30% V_N	50% S_N	50% V_N	70% S_N	70% V_N	90% S_N	90% V_N
1	1.0250	0.0270	1.0892	0.1331	1.1778	0.4091	1.3290	1.3800	1.6737	6.5196
2	1.0016	0.0016	1.0147	0.0178	1.0445	0.0777	1.1100	0.3552	1.3036	2.6437
3	1.0001	0.0001	1.0039	0.0043	1.0174	0.0261	1.0549	0.1485	1.1879	1.4875
4	1.0000	0.0000	1.0013	0.0013	1.0084	0.0113	1.0326	0.0772	1.1326	0.9665
5	1.0000	0.0000	1.0005	0.0005	1.0045	0.0056	1.0213	0.0455	1.0688	0.2930
10	1.0000	0.0000	1.0000	0.0000	1.0004	0.0004	1.0044	0.0065	1.0064	0.0292
20	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0009	1.0000	0.0105

表 6 分離連鎖法と線形法における平均成功アクセス回数の比較（上段は分離連鎖法，下段は線形法を示す）
Table 6 Comparisons of the average number of accesses between separate chaining and linear probing in a successful search for uniform probing.

Bucket size (b)	Load factor (α) 10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	1.0254	1.0516	1.0787	1.1066	1.1353	1.1649	1.1952	1.2263	1.2583	1.2745
	1.0425	1.1066	1.1875	1.2921	1.4323	1.6282	1.9176	2.3771	3.1797	3.9881
2	1.0017	1.0071	1.0166	1.0303	1.0484	1.0712	1.0986	1.1307	1.1674	1.1875
	1.0035	1.0169	1.0419	1.0818	1.1434	1.2394	1.3961	1.6736	2.2369	2.7651
3	1.0001	1.0011	1.0041	1.0103	1.0211	1.0375	1.0606	1.0912	1.1299	1.1523
	1.0004	1.0038	1.0132	1.0320	1.0656	1.1235	1.2260	1.4218	1.8597	2.3533
4	1.0000	1.0001	1.0009	1.0035	1.0095	1.0211	1.0401	1.0685	1.1079	1.1320
	1 0001	1.0009	1.0048	1.0146	1.0347	1.0732	1.1463	1.2953	1.6536	2.0531
5	1.0000	1.0000	1.0000	1.0009	1.0040	1.0117	1.0271	1.0530	1.0924	1.1178
	1.0000	1.0003	1.0019	1.0073	1.0200	1.0469	1.1019	1.2204	1.5236	1.9025
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0003	1.0139	1.0485	1.0767
	1.0000	1.0000	1.0000	1.0004	1.0023	1.0089	1.0257	1.0531	1.0752	1.0790
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0112	1.0393
	1.0000	1.0000	1.0000	1.0000	1.0001	1.0007	1.0028	1.0060	1.0084	1.0088

Note: An upper level of each row shows the average number of accesses in separate chaining.

るとき，あふれの処理としては，分離連鎖法や線形法を用いることが有効である．Knuth は文献1)で，これらの衝突処理間におけるアクセス回数の比較を行っているが，分離連鎖法を用いたときの評価方法は文献4)で指摘しているごとく，厳密な解析方法とはいえず，また，線形法における評価も前述したように近似値を与えるものである．

ここでは，探索頻度が一様なとき，分離連鎖法と線形法において，厳密な解析の下で評価した平均アクセス回数の数値例を**表6**に示す．

表6からわかるように，バケットの大きさが5までは，線形法の平均アクセス回数が分離連鎖法のそれに比べて大きく，とくに表占有率が大きいときには，その差は顕著である．しかし，バケットの大きさが10のときには，その差はほとんど認められなくなる．そしてバケットの大きさが20で，表占有率が90%ないし95%になると，逆に線形法の平均アクセス回数が，分離連鎖法のそれよりも小さくなることがわかる．

4. むすび

2次記憶を含む系での探索に，分散記憶法を適用し，ファイルへのアクセス回数を減少させる方法としては，複数個の見出しをひとまとめにして取り扱うバケット方式が用いられる．このとき，あふれの処理に

は，衝突処理のうち，分離連鎖法や線形法を用いることが有効である．

線形法については，Knuth によって，各見出しの探索頻度が一様であるという仮定のもとではあるが，平均アクセス回数を評価する表現式が導出されている．この評価式は簡潔に表現されてはあるが，アクセス回数の近似値を評価するものである．

本論文では，各見出しの探索頻度を考慮に入れ，しかも，アクセス回数をより厳密に評価する表現式を導き出した．次に，この評価式を用い，探索頻度に具体的な確率分布を与えたときのアクセス回数を評価した．とくに，探索頻度が一様な場合，Knuth の評価式と比較検討した．また，線形法と分離連鎖法でのアクセス回数を，厳密な解析のもとで評価して，比較することができた．

提案する評価式を用いることにより，バケット方式におけるアクセス回数を現実に即して評価することができるのみならずあふれの処理方法間での効率を厳密に比較することができる．

謝辞 本論文をまとめる際，城島邦行教授（熊本女子大学）および大島洋一教授（熊本大学工学部）に貴重なご意見をいただいた．ここに記して謝意を表す．

参 考 文 献

1) Knuth, D. E.: *The Art of Computer Programming*, Vol. 3, *Sorting and Searching*, pp. 534-538, Addison-Wesley, Reading, Mass. (1973).
2) Schay, G. and Spruth, W.: Analysis of File Addressing Method, *Comm. ACM*, Vol. 5, No. 8, pp. 459-462 (1962).
3) 中村，松山：見出しの探索頻度を考慮した探索路長の考察，情報処理学会論文誌，Vol. 24, No. 4, pp. 505-512 (1983).
4) 中村，城島，松山：分離連鎖法におけるバケット方式を用いたときのアクセス回数について，情報処理学会論文誌，Vol. 25, No. 1, pp. 141-149 (1984).

（昭和58年10月 4 日受付）
（昭和59年 7 月19日採録）