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†

中村良三††　城島邦行†††　松山公一††††

分散記憶法は2次記憶系を含むランダムアクセスのアドレスシングの問題に対しても有効に適用できる．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．Knuthは，このバケット方式で，各見出しの探索頻度が一様であるという仮定のもと，衝突処理のうち分離連鎖法，合併連鎖法，線形法について，すでにアクセス回数を評価する表現式を提案しているが，その導出前提に検討を要する問題がある．本論文では，分離連鎖法において，各見出しの探索頻度を考慮する観点から，アクセス回数の評価式を導出し，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，各見出しの探索頻度が一様である場合，Knuthの表現式と比較検討し，その相違点を指摘する．そして，その原因が導出前提の設定の違い，すなわち，確率変数の考え方の違いであることを示す．次に，適切なバケットサイズの算定が行えるように，あふれの起こる確率を評価できる表現式を導出し，具体的にあふれの起こる確率を考察する．

1. まえがき

言語処理系において名前表を高速に探索する方法として，分散記憶法が用いられているが，2次記憶系におけるランダムアクセスのアドレスシングの問題に対しても，分散記憶法は有効に適用できる．この問題を2次記憶系におけるファイルのアクセス回数の低減の問題としてとらえるとき，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考えられる．この方式では，ひとつのバケットにバケットサイズ分だけの見出しが格納されるので，同じ分散番地をもつ同族の見出しの数がバケットサイズ以上になったとき，はじめて，あふれが生ずることになる．このあふれの処理方法としては種々あるが，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方法が望ましい．この観点から，あふれの処理法としては，分離連鎖法や線形法が有効である．

このバケット方式で，Knuthは見出しの探索頻度を一様と仮定したもとで，アクセス回数を評価する表現式を，分離連鎖法，合併連鎖法および線形法の各技法について導き出している．そのなかで論じている分離連鎖法では，成功アクセス回数の評価式の導出前提には，成功探索路長を評価する表現式を導出したときの前提[2]を使用している．

しかし，この導出前提では，文献4)で指摘しているごとく，この成功探索路長の評価値は現実の現象に適合しなかった．そして，その原因が確率変数の設定の違いにあることが明示されている．

したがって，この前提を用いて導出されたアクセス回数の評価式にも同様な問題が起こると考えられる．

本論文では，各見出しの探索頻度を考慮する観点から，アクセス回数を確率変数としたモデルを作り，平均アクセス回数と分散を評価する表現式を導き出し，この評価式を用いて，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，各見出しの探索頻度が一様な場合，Knuthの評価式と比較検討する．

次に，適切なバケットサイズの算定が行えるように，あふれの起こる確率を考察し，その確率の評価式を導出し，あふれの起こる確率を評価する．

この導出過程では，分散表の大きさをM，バケットサイズをb，見出しの総数をN，表占有率をα $\left(\alpha=\frac{N}{Mb}\right)$ 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 $\rho_i(i=1,2,\cdots,N)$ 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．

成功時のアクセス回数については，次のような三つの頻度分布について論議するとともに数値例を示す．

a) 見出しが探索される確率が一様な場合：この

† On the Number of Accesses for the Buckets Storage Technique in Separate Chaining Method by RYOZO NAKAMURA (Kumamoto University), KUNIYUKI JOJIMA (Kumamoto Women's University) and KIMIKAZU MATSUYAMA (Kumamoto University).

†† 熊本大学電子計算機室
††† 熊本女子大学数学教室
†††† 熊本大学学長

とき，$\rho_i=\frac{1}{N}$ $(i=1,\cdots,N)$ となる．

b） 見出しが探索される確率が登録順序に従い半減する場合：このとき，$\rho_i=\frac{C}{2^{i-1}}(i=1,\cdots,N)$ となる．ただし，$C=\frac{1}{2-2^{1-N}}$ とする．

c） 見出しが探索される確率が登録順序に従い調和減少する場合：このとき，$\rho_i=\frac{C}{i}(i=1,\cdots,N)$ となる．ただし，$C=\frac{1}{H_N}$ で，$H_N=1+\frac{1}{2}+\cdots+\frac{1}{N}$ とする．

とくに，頻度が一様なa)の場合には，Knuth の表現式と比較検討する．

2. 提案する表現式

分離連鎖法では，同一の分散番地をもつ見出しはその分散番地を探索根とする一つのリストに連結し，その同族の見出しはバケットの先頭から順次登録する．しかし，バケットサイズを超えた見出しは，あふれ領域にひとつずつ格納し，順次連結する．

一方，探索するときには，基本的にはバケット単位にアクセスするが，あふれ領域に対しては見出し単位に行うものと考える．

このとき，バケットサイズを適当な大きさに選べば，あふれの起こる機会は，確率的にきわめて小さくすることができ，かつそれを数量的に把握することができる．

このバケット方式における平均成功アクセス回数の導出において，Knuth は，見出し単位に連結してリストを生成し，その探索路長を評価する表現式を導出したときの前提を使用している[2]．この導出前提では，すべての見出しを分配し，その任意の分配に対して，すべての見出しを探索したときの探索路長の総和を見出しの総数で割った，すなわち見出し一つ当りの探索路長を確率変数にとっているので，探索回数それ自身を確率変数にとった場合の探索路長に比べ，平均は若干大きめであるが，分散は著しく小さめになり，現実の現象に適合しないことが文献4) で指摘されている．したがって，この前提をバケット方式におけるアクセス回数の考察に使用すると，前述のような問題が起こると考えられる．

このようなことから，本論文では，アクセス回数自身を確率変数にとり，各見出しの探索頻度を考慮する観点から[4),5)]，その確率分布を求め，アクセス回数の平均，分散を評価する表現式を提案する．また，あふれの起こる確率を考察し，その確率を評価する表現式を導き出す．

2.1 アクセス回数

はじめに，アクセス回数を評価する表現式の導出過程で，次のような記号を導入する．

p_{Nk}：N個の見出しを大きさMの分散表に一様に分配するとき，同じ分散番地の見出しがk個になる確率．

この p_{Nk} は，任意のリストにk個の見出しが格納される確率で，次のように表現される．

$$p_{Nk}=\binom{N}{k}\left(\frac{1}{M}\right)^k\left(1-\frac{1}{M}\right)^{N-k},\ (k=0,1,\cdots,N) \tag{1}$$

このとき

$$\sum_{k=0}^{N}p_{Nk}=\sum_{k=0}^{N}\binom{N}{k}\left(\frac{1}{M}\right)^k\left(1-\frac{1}{M}\right)^{N-k}=1 \tag{2}$$

が成立する．

r_{kj}：k個の見出しからなる任意のリストで，先頭からj番目の見出しが探索される確率．

ここでは，同族の見出しは，登録される順序に従い，バケットの先頭から順々に挿入されると考えているので，N個の見出しのうち，i番目に登録された見出しが，k個の見出しからなるリストの先頭からj番目に配置される確率は $\binom{i-1}{j-1}\binom{N-i}{k-j}\Big/\binom{N-1}{k-1}$ となる[4)]．したがって，このリストの先頭からj番目に，すべての見出しが配置される可能性を考慮すると，確率 r_{kj} は次のように表現される．

$$r_{kj}=\sum_{i=1}^{N}\binom{i-1}{j-1}\binom{N-i}{k-j}\rho_i\Big/\binom{N-1}{k-1} \tag{3}$$

このとき，k個のすべての見出しが探索される確率は

$$\sum_{j=0}^{k}r_{kj}=1 \tag{4}$$

となる．ただし，

$$r_{k0}=\begin{cases}1 & (k=0)\\ 0 & (k>0)\end{cases}$$

q_{Nh}：任意のリストに連結されたバケットに通し番号1，あふれ領域の見出し単位に順次通し番号2, 3, …，と付けたとき，通し番号h番目がアクセスされる確率 $(h=0,1,\cdots,N-b+1)$．

探索は，あふれのない場合には，バケットを1回アクセスするのみであるが，あふれのある場合には，さらにあふれ領域に連結されている見出しをひとつずつ

アクセスすることになるから，先に導入した p_{Nk}, r_{kj} を用いて表現すると，q_{Nk} は次のようになる．

$$q_{Nk}=\begin{cases} p_{N0} & (h=0) \\ \sum_{i=1}^{b}\sum_{j=i}^{N} r_{ji}p_{Nj} & (h=1) \\ \sum_{j=h+b-1}^{N} r_{j\,h+b-1}p_{Nj} & (h>1) \end{cases} \tag{5}$$

このとき，

$$\sum_{h=0}^{N-b+1} q_{Nh}=1 \tag{6}$$

が成立する．

すなわち，

$$\begin{aligned}\sum_{h=0}^{N-b+1} q_{Nh}&=q_{N0}+q_{N1}+\sum_{h=2}^{N-b+1} q_{Nh}\\ &=p_{N0}+\sum_{i=1}^{b}\sum_{j=i}^{N} r_{ji}p_{Nj}\\ &\quad+\sum_{h=2}^{N-b+1}\sum_{j=h+b-1}^{N} r_{j\,h+b-1}p_{Nj}\\ &=p_{N0}+\sum_{i=1}^{N}\sum_{j=i}^{N} r_{ji}p_{Nj}\\ &\quad+\sum_{i=b+1}^{N}\sum_{j=i}^{N} r_{ji}p_{Nj}\\ &=p_{N0}+\sum_{i=1}^{N}\sum_{j=i}^{N} r_{ji}p_{Nj}\\ &=p_{N0}+\sum_{i=1}^{N}\sum_{j=1}^{i} r_{ij}p_{Ni}\\ &=p_{N0}+\sum_{i=1}^{N} p_{Ni}\\ &=1\end{aligned}$$

となる．

ところで，成功探索では，探すリストは少なくともひとつ以上の見出しが連結されているという条件を加味し，アクセス回数 h を確率変数にとれば，アクセス回数の平均，分散は前述の記号を用いて次のように表現される．

$$\begin{aligned}S_N&=\sum_{h=1}^{N-b+1} hq_{Nh}\Big/\sum_{k=1}^{N} p_{Nk}\\ &=\Big\{\sum_{i=1}^{b}\sum_{j=i}^{N} r_{ji}p_{Nj}\\ &\quad+\sum_{h=2}^{N-b+1} h\sum_{j=h+b-1}^{N} r_{j\,h+b-1}p_{Nj}\Big\}\\ &\quad\Big/\sum_{k=1}^{N} p_{Nk}\end{aligned} \tag{7}$$

$$V_N=\sum_{h=1}^{N-b+1} h^2q_{Nh}\Big/\sum_{k=1}^{N} p_{Nk}-S_N^2$$

$$\begin{aligned}&=\Big\{\sum_{i=1}^{b}\sum_{j=i}^{N} r_{ji}p_{Nj}\\ &\quad+\sum_{h=2}^{N-b+1} h^2\sum_{j=h+b-1}^{N} r_{j\,h+b-1}p_{Nj}\Big\}\\ &\quad\Big/\sum_{k=1}^{N} p_{Nk}-S_N^2\end{aligned} \tag{8}$$

不成功探索では，バケットを1回アクセスするか，あふれがあれば，あふれ領域のすべての見出しをアクセスすることになるので，そのアクセス回数の平均，分散は次のようになる．

$$\begin{aligned}\bar{S}_N&=\sum_{j=0}^{b} p_{Nj}+\sum_{j=b+1}^{N}(j-b+1)p_{Nj}\\ &=1+\sum_{j=b+1}^{N}(j-b)p_{Nj}\end{aligned} \tag{9}$$

$$\begin{aligned}\bar{V}_N&=\sum_{j=0}^{b} p_{Nj}+\sum_{j=b+1}^{N}(j-b+1)^2p_{Nj}-\bar{S}_N^2\\ &=\sum_{j=b+1}^{N}(j-b)^2p_{Nj}-\Big\{\sum_{j=b+1}^{N}(j-b)p_{Nj}\Big\}^2\end{aligned} \tag{10}$$

ところで，表の大きさ M や見出しの個数 N を実用的な大きさにとれば，2項分布の確率 p_{Nk} はポアソン分布で次のように近似できる．

$$\begin{aligned}p_{Nk}&=\binom{N}{k}\left(\frac{1}{M}\right)^k\left(1-\frac{1}{M}\right)^{N-k}\\ &\fallingdotseq e^{-\frac{N}{M}}\left(\frac{N}{M}\right)^k/k!\end{aligned} \tag{11}$$

ただし，e は自然対数の底を表す．

したがって，$\bar{S}_N, \bar{V}_N$ はこの(11)と表占有率 α を用いて，次のように表現できる．

$$\begin{aligned}\bar{S}_N&=1+\sum_{j=b+1}^{N}(j-b)p_{Nj}\\ &\fallingdotseq 1+\sum_{j=b+1}^{N}(j-b)e^{-\alpha b}(\alpha b)^j/j!\\ &=1+\alpha b t_b(\alpha)\end{aligned} \tag{12}$$

この(12)は Knuth の表現式に一致する．

$$\begin{aligned}\bar{V}_N&=\sum_{j=b+1}^{N}(j-b)^2p_{Nj}-\Big\{\sum_{j=b+1}^{N}(j-b)p_{Nj}\Big\}^2\\ &=2\sum_{j=b+1}^{N}\{j(j-1)-2j(b-1)+b(b-1)\}p_{Nj}\\ &\quad-\sum_{j=b+1}^{N}(j-b)p_{Nj}-\Big\{\sum_{j=b+1}^{N}(j-b)p_{Nj}\Big\}^2\\ &\fallingdotseq e^{-\alpha b}(\alpha b)^{b+1}b!^{-1}\{\alpha b-b+2\\ &\quad+(\alpha^2 b-2\alpha(b-1)+b-1)R(\alpha,b)\}\\ &\quad-\alpha b t_b(\alpha)-\{\alpha b t_b(\alpha)\}^2\end{aligned} \tag{13}$$

ただし，

$$t_b(\alpha)=e^{-\alpha b}\left(\frac{(\alpha b)^b}{(b+1)!}+\frac{2(\alpha b)^{b+1}}{(b+2)!}+\frac{3(\alpha b)^{b+2}}{(b+3)!}+\cdots\right) \quad (14)$$

$$R(\alpha,b)=\left(\frac{b}{b+1}+\frac{\alpha b^2}{(b+1)(b+2)}+\frac{\alpha^2 b^3}{(b+1)(b+2)(b+3)}+\cdots\right) \quad (15)$$

と定義する[2].

このとき，$t_b(\alpha)$ と $R(\alpha,b)$ の間には，次のような関係が成立する[2].

$$t_b(\alpha)=e^{-\alpha b}(\alpha b)^b b!^{-1}(1-(1-\alpha)R(\alpha,b)) \quad (16)$$

また，次の漸近公式

$$R_x(n)=1+\left(\frac{n}{n+1}\right)x+\left(\frac{n}{n+1}\right)\left(\frac{n}{n+2}\right)x^2+\cdots=\frac{1}{1-x}+\frac{x}{(1-x)^3 n}+O(n^{-2}) \quad (17)$$

(ただし，$x<1$)

が成立する[1]．したがって，(15)と(17)の間には次の関係が成立する．

$$R(\alpha,b)=\alpha^{-1}(R_\alpha(b)-1) \quad (18)$$

$$R(\alpha,b)=(1-\alpha)^{-1}-(1-\alpha)^{-3}b^{-1}+O(b^{-2}) \quad (19)$$

2.2 あふれの起こる確率

任意のリストで，見出しの個数が k 個になる確率 p_{Nk} と $k-1$ 個になる確率 $p_{N\,k-1}$ の比は(1)から次のようになる．

$$\frac{p_{Nk}}{p_{N\,k-1}}=1+\frac{(N+1)\frac{1}{M}-k}{k\left(1-\frac{1}{M}\right)} \quad (20)$$

このとき，p_{Nk} は $k<\frac{1}{M}(N+1)$ のとき単調増加し，$k>\frac{1}{M}(N+1)$ のとき単調減少するから，任意のリストで最も確からしい見出しの個数は $(N+1)\frac{1}{M}-1$ と $(N+1)\frac{1}{M}$ の間にある．また，(20)における比は k が増加するに従い単調に減少する．したがって $k\geq b+1$ ならば

$$\frac{p_{Nk}}{p_{N\,k-1}}\leq\frac{(N-b)\frac{1}{M}}{(b+1)\left(1-\frac{1}{M}\right)} \quad (21)$$

が成立するので，$k=b+1,b+2,\cdots,b+i$ とおいて，この i 個の不等式をかけ合わせると次の式が得られる．

$$\frac{p_{N\,b+i}}{p_{Nb}}\leq\left\{\frac{(N-b)\frac{1}{M}}{(b+1)\left(1-\frac{1}{M}\right)}\right\}^i \quad (22)$$

このとき，

$b>\frac{N}{M}$ ならば，$\frac{(N-b)\frac{1}{M}}{(b+1)\left(1-\frac{1}{M}\right)}<1$

であるから，(22)の右辺は収束幾何級数となり，次の不等式が成立する．

$$\sum_{i=1}^{N-b}p_{N\,b+i}\leq p_{Nb}\frac{N-b}{M(b+1)-(N+1)} \quad (23)$$

したがって，バケットサイズ b を $\frac{N}{M}$ より大きくとれば，あふれの起こる確率はたかだか

$$p_{Nb}\frac{N-b}{M(b+1)-(N+1)}$$

となる．

ここで，b を2項分布の中央の項，$(N+1)\frac{1}{M}$，すなわち，任意のリストで見出しが格納される数が最も確からしい値にとれば，p_{Nb} は $\left\{2\pi N\frac{1}{M}\left(1-\frac{1}{M}\right)\right\}^{-\frac{1}{2}}$ で近似できるので[3]，(23)は次のように表される．

$$\sum_{i=1}^{N-b}p_{N\,b+i}\leq\left(N-\frac{1}{M}(N+1)\right)(2\pi N(M-1))^{-\frac{1}{2}} \quad (24)$$

また，(23)の確率 p_{Nb} を(11)のポアソン分布で近似すれば，次のようになる．

$$\sum_{i=1}^{N-b}p_{N\,b+i}\leq e^{-\frac{N}{M}}\left(\frac{N}{M}\right)^b(b!)^{-1}\frac{N-b}{M(b+1)-(N+1)} \quad (25)$$

さらに，(25)を $\alpha b=\frac{N}{M}$ なる関係を用い，表占有率 $\alpha(\alpha<1)$ とバケットサイズ b を用いて，次のように表すことができる．

$$\sum_{i=1}^{N-b}p_{N\,b+i}<\frac{e^{-\alpha b}(\alpha b)^{b+1}}{b!(b+1-\alpha b)} \quad (26)$$

このとき，(26)と(25)の不等式の右辺の差は $e^{-\alpha b}(\alpha b)^b(b!)^{-1}\{(1-\alpha)(b^2+b)/(b(1-\alpha)+1)(Mb(1-\alpha)+M-1)\}$ となるが，実用上無視できる．

したがって，(25)，(26)は，バケットサイズ b $\left(b>\frac{N}{M}\right)$ を設定したとき，あふれの起こる確率の目安を与える簡潔に表現式となりえる．

表現式(25)を用いた数値例を**表1**に示す．この数値

表 1 バケットサイズに伴うあふれの起こる確率
Table 1 Probability of overflows according to the increase bucket size. Size of table M=100.

Bucket size (b)	Number of identifiers (N)					
	50		100		150	
	P_1	P_2	P_1	P_2	P_1	P_2
1	0.08943	0.09973	0.26423		0.44301	
2	0.01381	0.01461	0.07937	0.09058	0.19052	0.24933
3	0.00159	0.00170	0.01837	0.01989	0.06469	0.07409
4	0.00014	0.00016	0.00343	0.00368	0.01799	0.01968
5	0.00001	0.00001	0.00053	0.00058	0.00421	0.00455
10	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
20	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

表 2 表占有率およびバケットサイズに伴うあふれの起こる確率
Table 2 Probability of overflows according to the increase of the bucket size and load factor.

Bucket size (b)	Load factor (α)									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	0.0047	0.0182	0.0392	0.0670	0.1011	0.1411	0.1871	0.2396	0.2994	0.3324
2	0.0011	0.0082	0.0247	0.0523	0.0919	0.1445	0.2114	0.2953	0.4016	0.4663
3	0.0002	0.0034	0.0143	0.0372	0.0753	0.1314	0.2089	0.3135	0.4579	0.5531
4	0.0000	0.0014	0.0082	0.0259	0.0601	0.1157	0.1982	0.3166	0.4917	0.6154
5	0.0000	0.0006	0.0047	0.0180	0.0477	0.1008	0.1850	0.3125	0.5124	0.6624
10	0.0000	0.0000	0.0003	0.0030	0.0151	0.0495	0.1242	0.2647	0.5336	0.7822
20	0.0000	0.0000	0.0000	0.0000	0.0017	0.0129	0.0574	0.1797	0.4808	0.8258

例では，表の大きさ M を 100，見出しの個数 N を 50 (50) 150 とし，バケットサイズ b の大きさに伴うあふれの起こる確率を，$\sum_{i=1}^{N-b} p_{N\,b+i}$ なる式から逐一求めた場合の確率 P_1 と提案する表現式(25)から近似的に求めた場合の確率 P_2 について，対比して示す．

表1からわかるように，提案する表現式(25)を用いて，あふれの起こる確率を実用的な精度で求めることができる．

さらに，表現式(26)を用いれば，あふれの起こる確率を表占有率とバケットサイズのみによって把握することができる．この数値例を**表 2** に示す．

3. 比較検討

ここでは，見出しの探索頻度が一様な場合および登録順序に従い半減する場合と調和減少する場合について，アクセス回数を検討する．とくに，一様な場合には，Knuth の表現式と比較検討し，その相違点を論議する．

3.1 探索頻度が一様な場合

探索頻度が一様という仮定から，$\rho_i=\frac{1}{N}$ $(i=1,\cdots,N)$ となり，r_{kj} は(3)より次のようになる．

$$\begin{aligned} r_{kj}&=\sum_{i=1}^{N}\binom{i-1}{j-1}\binom{N-i}{k-j}\rho_i\Big/\binom{N-1}{k-1}\\ &=\frac{1}{N}\sum_{i=1}^{N}\binom{i-1}{j-1}\binom{N-i}{k-j}\Big/\binom{N-1}{k-1}\\ &=\binom{N}{k}\Big/N\binom{N-1}{k-1}\\ &=\frac{1}{k} \end{aligned} \tag{27}$$

したがって，成功アクセス回数の平均，分散は(7)，(8)から次のようになる．

$$\begin{aligned} S_N&=\left\{\sum_{i=1}^{b}\sum_{j=i}^{N}\frac{1}{j}p_{Nj}+\sum_{h=2}^{N-b+1}h\sum_{j=h+b-1}^{N}\frac{1}{j}p_{Nj}\right\}\Big/\sum_{k=1}^{N}p_{Nk}\\ &=\left\{p_{N1}+p_{N2}+\cdots+p_{NN}+\frac{p_{N\,b+1}}{b+1}+\cdots+\frac{(N-b)(N-b+1)p_{NN}}{2N}\right\}\Big/\sum_{k=1}^{N}p_{Nk}\\ &=1+\sum_{k=b+1}^{N}\frac{(k-b)(k-b+1)}{2k}p_{Nk}\Big/\sum_{k=1}^{N}p_{Nk} \end{aligned} \tag{28}$$

$$V_N=\left\{\sum_{i=1}^{b}\sum_{j=i}^{N}\frac{1}{j}p_{Nj}+\sum_{h=2}^{N-b+1}h^2\sum_{j=h+b-1}^{N}\frac{1}{j}p_{Nj}\right\}\Bigg/\sum_{k=1}^{N}p_{Nk}-S_N^2$$
$$=1+\sum_{k=b+1}^{N}\frac{(k-b+1)(k-b+2)(2(k-b)+3)}{6k}*$$
$$*\frac{-6(k-b)-6}{}p_{Nk}\Bigg/\sum_{k=1}^{N}p_{Nk}-S_N^2 \qquad (29)$$

ここで，後述する Knuth の表現式(38)と表現形式を対比させるため，この S_N, V_N を(11)のポアソン分布で近似し，表占有率 α，バケットサイズ b および関数 $R(\alpha,b)$ で表すと次のようになる．

$$S_N=1+\left\{\sum_{k=b+1}^{N}kp_{Nk}+(1-2b)\times\sum_{k=b+1}^{N}p_{Nk}+b(b-1)\sum_{k=b+1}^{N}\frac{p_{Nk}}{k}\right\}\Bigg/2\sum_{k=1}^{N}p_{Nk} \qquad (30)$$

この式の{…}内の第1項 $\sum_{k=b+1}^{N}kp_{Nk}$ は(11)を用いて近似すると次のようになる．

$$\sum_{k=b+1}^{N}kp_{Nk}\fallingdotseq\sum_{k=b+1}^{N}\frac{e^{-\alpha b}(\alpha b)^k}{(k-1)!}=e^{-\alpha b}(\alpha b)^{b+1}b!^{-1}\{1+\alpha R(\alpha,b)\} \qquad (31)$$

第2項の $\sum_{k=b+1}^{N}p_{Nk}$ についても同様に

$$\sum_{k=b+1}^{N}p_{Nk}\fallingdotseq\sum_{k=b+1}^{N}\frac{e^{-\alpha b}(\alpha b)^k}{k!}=e^{-\alpha b}(\alpha b)^{b+1}R(\alpha,b)(bb!)^{-1} \qquad (32)$$

となる．

また，第3項の $\sum_{k=b+1}^{N}p_{Nk}/k$ を $\sum_{k=b+1}^{N}p_{Nk}/(k+1)$ で近似すれば，そのときの誤差は $\sum_{k=b+1}^{N}p_{Nk}/k(k+1)$ となる．しかし，バケットサイズ b を適当な大きさにとれば，あふれの起こる確率 $\sum_{k=b+1}^{N}p_{Nk}$ は小さな値になり，この誤差も十分小さな値とみなせる．

よって，第3項の $\sum_{k=b+1}^{N}p_{Nk}/k$ は次のようになる．

$$\sum_{k=b+1}^{N}\frac{p_{Nk}}{k}\fallingdotseq\sum_{k=b+1}^{N}p_{Nk}/(k+1)\fallingdotseq\sum_{k=b+1}^{N}e^{-\alpha b}(\alpha b)^k/(k+1)!=e^{-\alpha b}(\alpha b)^{b+1}b!^{-1}\{(b+1)R(\alpha,b)-b\}/\alpha b^2(b+1) \qquad (33)$$

(30) に (31)，(32)，(33) を代入し，

$\sum_{k=1}^{N}p_{Nk}=1-e^{-\alpha b}$ [4]

を用いれば，S_N は次のようになる．

$$S_N=1+\frac{1}{2}e^{-\alpha b}(\alpha b)^{b+1}(1-e^{-\alpha b})^{-1}(b!)^{-1}\times\{1-(b-1)\alpha^{-1}(b+1)^{-1}+R(\alpha,b)(\alpha+\alpha^{-1}+b^{-1}-(\alpha b)^{-1}-2)\} \qquad (34)$$

同様に，V_N も次のようになる．

$$V_N=1+\left\{2\sum_{k=b+1}^{N}k(k-1)p_{Nk}+(11-6b)\times\sum_{k=b+1}^{N}kp_{Nk}+(6b^2-18b+7)\times\sum_{k=b+1}^{N}p_{Nk}-b(2b-7)(b-1)\sum_{k=b+1}^{N}p_{Nk}/k\right\}\Bigg/6\sum_{k=1}^{N}p_{Nk}-S_N^2 \qquad (35)$$

このとき，(35)の{…}内の第1項 $\sum_{k=b+1}^{N}k(k-1)p_{Nk}$ を(11)で近似すると次のようになる．

$$\sum_{k=b+1}^{N}k(k-1)p_{Nk}\fallingdotseq\sum_{k=b+1}^{N}e^{-\alpha b}(\alpha b)^k/(k-2)!=e^{-\alpha b}(\alpha b)^{b+1}b!^{-1}\{\alpha+\alpha b+\alpha^2bR(\alpha,b)\} \qquad (36)$$

また，第2項の $\sum_{k=b+1}^{N}kp_{Nk}$，第3項の $\sum_{k=b+1}^{N}p_{Nk}$，第4項の $\sum_{k=b+1}^{N}\frac{p_{Nk}}{k}$ は，それぞれ(31)，(32)，(33)で表現されるので，結局，(35)の分散は次のようになる．

$$V_N=1+\frac{1}{6}e^{-\alpha b}(\alpha b)^{b+1}(1-e^{-\alpha b})^{-1}(b!)^{-1}\{2\alpha b-4b+11+(b-1)(2b-7)\alpha^{-1}(b+1)^{-1}+R(\alpha,b)(2\alpha^2b+(11-6b)\alpha+6b-18+7b^{-1}-(b-1)(2b-7)(\alpha b)^{-1})\}-S_N^2 \qquad (37)$$

他方，Knuth は，各見出しの探索頻度が一様という条件のもとで，平均成功アクセス回数 S_N を次のように表している[2]．

$$S_N=1+\frac{M}{N}\sum_{k>b}^{N}\binom{k-b+1}{2}p_{Nk}\fallingdotseq1+\frac{1}{2}e^{-\alpha b}(\alpha b)^b b!^{-1}(\alpha b-b+2+(\alpha^2b-2\alpha(b-1)+b-1)R(\alpha,b))=1+(1-b(1-\alpha)/2)t_b(\alpha)+e^{-\alpha b}(\alpha b)^bR(\alpha,b)/2b! \qquad (38)$$

ここでの関数 $t_b(\alpha), R(\alpha,b)$ は(14)，(15)で定義されている．

分散については，それを評価する表現式は導出され

てはいないが，(38)の S_N を導出した前提にもとづいて，V_N を導出するとすれば，次のようになるものと思われる．

$$
\begin{aligned}
V_N=&\frac{\Sigma\binom{N}{k_1,\cdots,k_M}}{M^N}\\
&\times\left\{\frac{\binom{k_1-b+1}{2}+\cdots+\binom{k_M-b+1}{2}}{N}\right\}^2\\
&-(S_N-1)^2\\
=&\frac{1}{M^NN^2}\Bigg\{M(M-1)\\
&\times\Sigma\binom{N}{k_1,\cdots,k_M}\binom{k_1-b+1}{2}\\
&\times\binom{k_2-b+1}{2}(M-2)^{N-k_1-k_2}\\
&+M\Sigma\binom{N}{k_1,\cdots,k_M}\binom{k_1-b+1}{2}^2(M-1)^{N-k_1}\Bigg\}\\
&-(S_N-1)^2\\
=&\frac{1}{M^NN^2}\Bigg\{M(M-1)\sum_{k_1>b}\binom{N}{k_1}\binom{k_1-b+1}{2}\\
&\times\sum_{k_2>b}\binom{N-k_1}{k_2}\binom{k_2-b+1}{2}(M-2)^{N-k_1-k_2}\\
&+M\sum_{k_1>b}\binom{N}{k_1}\binom{k_1-b+1}{2}^2(M-1)^{N-k_1}\Bigg\}\\
&-(S_N-1)^2 \qquad (39)
\end{aligned}
$$

上の式で示されるように，Knuth の前提では，分散 V_N を簡潔に表現することはきわめてむずかしい．

このように，提案する表現式と Knuth の表現式の違いは，その導出前提の違い，すなわち確率変数の考え方の違いにある．提案する表現式を用いれば，成功アクセス回数の平均，分散とも表占有率 α，バケットサイズ b および関数 $R(\alpha, b)$ で簡潔に表現することができる．

ここで，提案する表現式(34)から求めた成功アクセス回数の平均と Knuth が提案している表現式(38)から求まる平均アクセス回数の数値例を**表3**，**表4**に示す．また，提案する表現式(37)から求めた分散を**表5**に示す．

この数値例からわかるように，Knuth の表現式から求めた平均アクセス回数は提案する表現式から求めた平均アクセス回数に比べ若干大きめになる．

表 3 提案する表現式を用いたときの平均成功アクセス回数
Table 3 Average accesses in a successful search by the proposed formula.

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
2	1.0017	1.0071	1.0166	1.0303	1.0484	1.0712	1.0986	1.1307	1.1674	1.1875
3	1.0001	1.0011	1.0041	1.0103	1.0211	1.0375	1.0606	1.0912	1.1299	1.1523
4	1.0000	1.0001	1.0009	1.0035	1.0095	1.0211	1.0401	1.0685	1.1079	1.1320
5	1.0000	1.0000	1.0000	1.0009	1.0040	1.0117	1.0271	1.0530	1.0924	1.1178
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0003	1.0139	1.0485	1.0767
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0112	1.0393

表 4 Kunth の表現式を用いたときの平均成功アクセス回数
Table 4 Average accesses in a successful search by Knuth's fomula.

Bucket size (b)	Load factor (α)									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	1.0500	1.1000	1.1500	1.2000	1.2500	1.3000	1.3500	1.4000	1.4500	1.4750
2	1.0063	1.0242	1.0519	1.0883	1.1321	1.1823	1.2381	1.2988	1.3637	1.3976
3	1.0010	1.0071	1.0215	1.0458	1.0806	1.1259	1.1812	1.2459	1.3192	1.3589
4	1.0002	1.0023	1.0097	1.0257	1.0526	1.0922	1.1449	1.2111	1.2899	1.3340
5	1.0000	1.0008	1.0046	1.0151	1.0358	1.0699	1.1195	1.1856	1.2684	1.3159
10	1.0000	1.0000	1.0002	1.0015	1.0070	1.0226	1.0559	1.1151	1.2063	1.2653
20	1.0000	1.0000	1.0000	1.0000	1.0005	1.0038	1.0182	1.0597	1.1502	1.2206

表 5 提案する表現式を用いたときの成功アクセス回数の分散
Table 5 Variance of accesses in a successful search by the proposed formula.

Bucket size (b)	Load factor (α)									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
1	0.0259	0.0536	0.0833	0.1148	0.1483	0.1838	0.2212	0.2607	0.3022	0.3236
2	0.0029	0.0123	0.0292	0.0543	0.0885	0.1324	0.1866	0.2515	0.3274	0.3695
3	0.0005	0.0044	0.0148	0.0350	0.0681	0.1169	0.1842	0.2722	0.3826	0.4467
4	0.0001	0.0019	0.0089	0.0261	0.0587	0.1125	0.1929	0.3047	0.4521	0.5401
5	0.0000	0.0009	0.0057	0.0205	0.0529	0.1118	0.2055	0.3421	0.5284	0.6420
10	0.0000	0.0000	0.0007	0.0067	0.0329	0.1062	0.2597	0.5251	0.9283	1.1884
20	0.0000	0.0000	0.0000	0.0005	0.0084	0.0625	0.2638	0.7559	1.6568	2.2816

この現象は，文献4)で指摘した平均探索路長の場合と同様な傾向を示す．また，提案する表現式から求めた分散値は，表占有率およびバケットサイズに伴うあふれの起こる確率の数値例，表2と対比して考察すれば，その現象をよく把握できるであろう．

3.2 探索頻度に確率分布を与えた場合

1) 探索頻度が登録順序に従い半減するとき

1章b)から，$\rho_i=1/(2^{i-1})(2-2^{1-N})$, $(i=1,\cdots,N)$ となる．このとき，(3)から r_{kj} は次のようになる．

$$r_{kj}=\frac{1}{1-2^{-N}}\sum_{i=1}^{N}\binom{i-1}{j-1}\binom{N-i}{k-j}\frac{1}{2^i}\Big/\binom{N-1}{k-1} \tag{40}$$

2) 探索頻度が登録順序に従い調和減少するとき

1章c)から，$\rho_i=\dfrac{1}{iH_N}$, $(i=1,\cdots,N)$ となる．したがって，(3)から r_{kj} は次のようになる．

表 6 見出しの探索頻度が一様な場合の成功アクセス回数
Table 6 The number of accesses in a successful search for uniform probing.

Bucket size (b)	Number of identifiers (N)					
	50		100		150	
	S_N	V_N	S_N	V_N	S_N	V_N
1	1.1329	0.1447	1.2886	0.3409	1.4633	0.5878
2	1.1013	0.0158	1.0567	0.0757	1.1297	0.1925
3	1.0001	0.0001	1.0014	0.0017	1.0067	0.0097
4	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
10	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

表 7 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 7 The number of accesses in a successful search for the reduction by half.

Bucket size (b)	Number of identifiers (N)					
	50		100		150	
	S_N	V_N	S_N	V_N	S_N	V_N
1	1.0054	0.0054	1.0058	0.0059	1.0062	0.0063
2	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
3	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
5	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
20	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000

表 8 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 8 The number of accesses in a successful search for the harmonic reduction.

Bucket size (b)	Number of identifiers (N)					
	50		100		150	
	S_N	V_N	S_N	V_N	S_N	V_N
1	1.0548	0.0613	1.1066	0.1337	1.1606	0.2246
2	1.0043	0.0051	1.0164	0.0218	1.0353	0.0532
3	1.0000	0.0000	1.0003	0.0004	1.0015	0.0022
4	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
10	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

$$r_{kj}=\frac{1}{H_N}\sum_{i=1}^{N}\binom{i-1}{j-1}\binom{N-i}{k-j}\frac{1}{j}\Big/\binom{N-1}{k-1} \qquad (41)$$

頻度を考慮した場合，平均成功アクセス回数および分散は，(40)，(41)を(7)，(8)に，それぞれ代入することによって評価される．

ここでは，数値例として，表の大きさMを 100，見出しの個数Nを 50(50)100 とし，バケットサイズbの大きさに伴う成功アクセス回数の平均，分散を，見出しの探索頻度が一様な場合，登録順序に従い半減する場合および調和減少する場合を**表 6**，**表 7** および**表 8** に示す．

4. むすび

分散記憶法が 2 次記憶を含む系まで適用されることを考慮すれば，バケット方式を用いたときのアクセス回数を具体的に考察することは意義があると考える．この観点から，本稿では，あふれの処理として有効である分離連鎖法を用い，各見出しの探索頻度を考慮したアクセス回数の評価式を求め，この評価式を用いて，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した．

とくに，探索頻度が一様な場合，Knuth の表現式と比較検討を行った．その結果，提案する表現式が，アクセス回数の解析方法として，見通しがよい上に，現実の現象をより適切に評価できると考える．

また，あふれの起こる確率を簡潔に，かつ十分な精度で評価できる表現式を導き出しているので，適切なバケットサイズの算定が容易に行える．

参考文献

1) Knuth, D. E.: *The Art of Computer Programming*, Vol. 1, *Fundamental Algorithms*, pp. 112-120, Addison-Wesley, Reading, Mass. (1973).
2) Knuth, D. E.: *The Art of Computer Programming*, Vol. 3, *Sorting and Searching*, pp. 534-538, Addison-Wesley, Reading, Mass. (1973).
3) Feller, W.: *Introduction to Probability and Its Application*, Vol. 1, pp. 139-141, John Wiley & Son, New York (1957).
4) 中村，松山：分散記憶法における探索頻度を考慮した探索距長とその評価，情報処理学会論文誌，Vol. 24, No. 1, pp. 125-130 (1983).
5) 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).

(昭和 58 年 5 月 13 日受付)
(昭和 58 年 7 月 19 日採録)