ショートノート

k 元数列においてあるパターンが初めて出現する位置について†

仙　波　一　郎††

k 元数列（k 種の相異なる記号を重複を許して並べた記号列）において，r ケタのパターン（r 個の記号を並べたもの）が初めて出現する位置 n（k 元数列を左から右へ走査していって，n 番目の記号を走査し終ったところで，パターンが現われるということ）の平均と分散を，k 種の記号が一様に現われるという仮定のもとでもとめた．

1. はじめに

Liu[1] によって，2元数列（2種の記号を重複を許して並べた記号列）における，パターン（1個以上の記号を並べたもの）の出現に関する種々の問題が解析されている．

この論文では，一般に k 元数列（k 種の記号を重複を許して並べた記号列）において，r ケタのパターン（r 個の記号を並べたもの）が，初めて n 番目に出現する（k 元数列を左から右へ走査していって，n 番目の記号を走査し終ったところで，パターンがあらわれるということ）場合の数 a_n が満たす母関数 $A(x)$ をもとめ，k 種の記号が一様に出現する仮定のもとで，位置 n の平均と分散をもとめた．

文献 2) には，パターンを固定しない場合についての統計的解析がなされている．

2. 定義

● 与えられた r ケタのパターン $P=P_1P_2\cdots P_r$ に対して ε_i $(0\leqq i\leqq r)$ をつぎのように定義する．

$0\leqq i\leqq r-1$ の場合

$$\varepsilon_i=\begin{cases}1 & \text{if } P_j=P_{j-i}(i+1\leqq j\leqq r)\\ 0 & \text{otherwise}\end{cases}$$

すなわち，パターン P の $i+1$ 番目から，r 番目までが，パターン P の1番目から，$r-i$ 番目までと一致するならば，$\varepsilon_i=1$，そうでないならば $\varepsilon_i=0$ とする．

$\varepsilon_r=1$ と定める．ε_0 はつねに1である．

● 母関数 $A(x)$ は $\sum_{i=0}^{\infty} a_i x^i$ とする．

明らかに $a_i=0$ $(0\leqq i\leqq r-1)$ である．

● 母関数 $B(x)$ は $\sum_{i=0}^{\infty} b_i x^i$ とする．

ただし $b_i=\varepsilon_i$ $(0\leqq i\leqq r-1)$

$b_i=k^{i-r}$ $(r\leqq i)$

とする．

したがって $B(x)=\sum_{i=0}^{r-1}\varepsilon_i x^i+x^r/(1-kx)$

となる．

3. 母関数 $A(x)$

a_n に関してつぎの性質がなりたつ．

[性質 1]

$n\geqq 2r$ について

$$k^{n-r}=\sum_{i=r}^{n-r}k^{n-r-i}a_i+\sum_{i=n-r+1}^{n}\varepsilon_{n-i}a_i \qquad (1)$$

（証明）　n ケタの k 元数列で最後の r ケタが，パターン P であるものは k^{n-r} 個ある．

つぎに，パターン P が初めて出現する場所 i によりこれらの k 元数列を分類すると，

● 場合 1　$r\leqq i\leqq n-r$ のとき．

$i+1$ 番目から $n-r$ 番目まで k 種の記号のどれでもよいので $k^{n-r-i}a_i$ 個．

● 場合 2　$n-r+1\leqq i\leqq n$ のとき．

$\varepsilon_{n-i}=1$ のとき a_i 個．$\varepsilon_{n-i}=0$ のとき 0個よりまとめると $\varepsilon_{n-i}a_i$ 個．

以上合計すれば(1)を得る．（終）

† On a Location of the First Occurrence of a String "Pattern" in a k-ary Sequence by ICHIRO SENBA (Department of Pure and Applied Sciences, College of General Education, University of Tokyo).

†† 東京大学教養学部基礎科学科

［性質 2］

$r \leqq n \leqq 2r-1$ について

$$k^{n-r}=\sum_{i=r}^{n} \varepsilon_{n-i} a_i \tag{2}$$

（証明）性質1の場合2と同様に考えればよい．

ただし $r \leqq i \leqq n$.　（終）

ここで $k^{n-r-i}=b_{n-i},\ \varepsilon_{n-i}=b_{n-i}$

$a_i=0\ (0 \leqq i \leqq r-1)$ と性質 1，2 を考えると，(1)，(2)はつぎのようにまとめられる．

$$k^{n-r}=\sum_{i=0}^{n} b_{n-i} a_i \quad (n \geqq r) \tag{3}$$

よって

$$\sum_{n=r}^{\infty} k^{n-r} x^n=\sum_{n=r}^{\infty}(b_n a_0+\cdots+b_0 a_n) x^n$$

$$\begin{aligned} x^r/(1-kx) &= A(x)B(x)-(b_0a_0) \\ &\quad -(b_1a_0+b_0a_1)x \\ &\quad \vdots \\ &\quad -(b_{r-1}a_0+\cdots+b_0a_{r-1})x^{r-1} \end{aligned}$$

上式の右辺は $A(x)B(x)$ 以外0のになるので整理してまとめるとつぎのようになる．

$$A(x)=\frac{x^r}{\sum_{i=0}^{r} \varepsilon_i x^i-k \sum_{i=0}^{r-1} \varepsilon_i x^{i+1}} \tag{4}$$

4. 平均と分散

k 種の記号が，一様に出現するとすると，k 元数列の i 番目でパターン P が初めて出現する確率は a_i/k^i である．

平均を L_r，分散を V_r とすると

$$L_r=\sum_{i=0}^{\infty} i a_i/k^i$$

$$V_r=\sum_{i=0}^{\infty}(i-L_r)^2 a_i/k^i$$

となる．つぎの定理がなりたつ．

［定理］

$$L_r=\sum_{i=0}^{r-1} \varepsilon_i k^{r-i} \tag{5}$$

$$V_r=L_r^2-(2r-1)L_r+2\sum_{i=0}^{r-1} i\varepsilon_i k^{r-i} \tag{6}$$

（証明）

$\sum_{i=0}^{\infty}(a_i/k^i)=1$ より $A(1/k)=1$ がなりたつ．L_r，V_r をもとめるのに母関数が利用できる．

$$L_r=\frac{1}{k}A'\left(\frac{1}{k}\right)$$

$$\begin{aligned} V_r &= \sum_{i=0}^{\infty} i^2 \frac{a_i}{k^i}-L_r^2 \\ &= \frac{1}{k}A'\left(\frac{1}{k}\right)+\frac{1}{k^2}A''\left(\frac{1}{k}\right)-L_r^2 \end{aligned}$$

となるので，$A'(1/k)$，$A''(1/k)$ を計算して代入すれば，(5)，(6)を得る．　（終）

おわりに，いつも貴重な助言をいただいている清水留三郎助教授に深く感謝致します．

参 考 文 献

1) Liu, C. L.：組合せ数学入門 I，伊理正夫訳，pp. 61-99，共立全書，東京（1972）．
2) 花田孝郎：パターン・マッチングのアルゴリズムについて，第 20 回プログラミング・シンポジウム報告集，pp. 19-29（1979）．

（昭和54年3月16日受付）
（昭和54年4月19日採録）