〈論　文〉

ある種の並べ替え算法における分布について

土屋　高宏
中村　永友

1　はじめに

本論文で扱う並べ替え(整列，ソーティング)算法は，バケットソートと呼ばれる計算時間が $O(n)$で高速に行えるものである。これは並べ替えをしたいデータの取りうる値が n 通りあるとき，あらかじめ n 個のバケツを用意しておき，あるいは動的にバケツを増やしていきながら，各々の数字と対応するバケツにデータを入れていく方法である。本論文は，n 個の連続した数字があり，あらかじめ用意するバケツ数が決められておらず，その数がデータの初期状態に依存する変形バケットソートを考え，そのバケツ数がどのような確率分布になるのかを考察する。

2 節では変形バケットソートのアルゴリズムについて述べ，3 節は確率分布を導出するための基本的な考え方について述べる。4 節は確率分布のモーメントを求め，5 節では正規分布への近似を試みる。

2　変形バケットソートのアルゴリズム

簡単のため数字の書かれたカードを例に説明する。

1 から n までの数字が書かれたよくシャッフルされている n 枚のカードが手元にある。小さい方から並べ替えを行うために，次の**手順**で行う。

1. 一番上のカードの数字が k のとき，数字 $k+1$ のカードがすでにテーブルに置かれていたらその上にカード k を載せる。
2. もしテーブル上に $k+1$ が無ければ，どのカードの上にも載せず，テーブルにカード k を置く。
3. 1 と 2 を手元のカードが無くなるまで続ける。
4. テーブルの上に m 個のカードの束ができる。$(1 \leq m \leq n)$
5. 各束の一番上のカードの数字を小さい順にまとめれば，並べ替えが完了する。

通常のバケットソートとの相違点は，(1) 手元の n 枚のカードが 1 から n までの重複がないという前提がある点，(2) バケツ数がいくつになるかわからないという点，(3) 数字 $k+1$ がすでに置かれていれば k を $k+1$ に載せる点，である。この方法による並べ替えでは，手順の 4 の段階で最悪の場合は n 個，最善の場合は 1 個のカードの束ができることになる。

このとき，カードをテーブルに置き終わった時点におけるカードの束の数（バケツ数に対応）に関する確率分布がどのようなものになるのかを議論する。

3　バケット数の確率分布

本節では関心のある確率分布の導出の基本的な考え方を示し，それに基づいて漸化式により確率分布を表現する。

3.1　基本的な考え方

$n=2$ のとき，カードの出方は$(1,2)$，$(2,1)$の 2 通りである。ただし，記号(a,b)はカードが a，b の順に現れるものとする。並べ替えの手順からカードの束の数は$(1,2)$のときは 2，$(2,1)$のときは 1 となることがわかる。次に，$n=3$ のとき，カードの出方は$(1,2,3)$，$(1,3,2)$，$(2,1,3)$，$(2,3,1)$，$(3,1,2)$，$(3,2,1)$の 6 通りである。ここでも，記号(a,b,c)はカードが a，b，c の順に現れるものとする。カードの束の数は$(1,2,3)$のときは 3，$(1,3,2)$，$(2,1,3)$，$(2,3,1)$，$(3,1,2)$のときは 2，$(3,2,1)$のときは 1 となる。さらに，$n=4$ のとき，カードの出方は 24$(=4!)$通りあり，それぞれのカードの順番と束の数は以下の通りである。

$$
\begin{cases}
(1,2,3,4)\cdots 4 & (2,1,3,4)\cdots 3 & (3,1,2,4)\cdots 3 & (4,1,2,3)\cdots 3 \\
(1,2,4,3)\cdots 3 & (2,1,4,3)\cdots 2 & (3,1,4,2)\cdots 3 & (4,1,3,2)\cdots 2 \\
(1,3,2,4)\cdots 3 & (2,3,1,4)\cdots 3 & (3,2,1,4)\cdots 2 & (4,2,1,3)\cdots 2 \\
(1,3,4,2)\cdots 3 & (2,3,4,1)\cdots 3 & (3,2,4,1)\cdots 2 & (4,2,3,1)\cdots 2 \\
(1,4,2,3)\cdots 3 & (2,4,1,3)\cdots 2 & (3,4,1,2)\cdots 3 & (4,3,1,2)\cdots 2 \\
(1,4,3,2)\cdots 2 & (2,4,3,1)\cdots 2 & (3,4,2,1)\cdots 2 & (4,3,2,1)\cdots 1
\end{cases}
$$

一般に n 枚のカードの束の数を求めるためには，$n-1$ 枚のときと n 枚のときのカードの出方と束の数の関係を考える必要がある。いま，$n=2$ と $n=3$ のときにこの関係をみる。

$n=3$ のカードの出方は，全部で 6 通りあるが，これを $n=2$ の場合から次のように考えることができる。$n=2$ の$(1,2)$において，3 のカードは$(a\ 1\ b\ 2\ c)$の 3 個のスペース a，b，c のいずれかに現れる（入る）ので 3 通りあり，$(2,1)$において，3 のカードは$(d\ 2\ e\ 1\ f)$の 3 個のスペース d，e，f のいずれかに入るので 3 通りあると考えると合計で 6 通りになる。

もし，3が$(a\ 1\ b\ 2\ c)$のaまたはbに入るとき，$(3,1,2)$または$(1,3,2)$となり，束の数はそれぞれ2であるのでその数は不変である。また，$(d\ 2\ e\ 1\ f)$のdに入るとき，$(3,2,1)$となり，束の数は1であるのでこの場合も不変である。一方，3が$(a\ 1\ b\ 2\ c)$のcに入るとき，$(1,2,3)$となり，束の数は3になるので1増えることになる。また，$(d\ 2\ e\ 1\ f)$のeまたはfに入るとき，$(2,3,1)$または$(2,1,3)$となり，束の数はそれぞれ2になるのでこの場合も1増えることになる。このように見ていくと，3が2の左側にあるときは束の数は不変となり，3が2の右側にあるときは束の数は1増えることがわかる。このことは一般に成り立ち，nが$n-1$の左側にあるときは束の数は不変となり，nが$n-1$の右側にあるときは束の数は1増えることがわかる。

3.2　2個の束の組合せ

いま，$n=3,4,5$のときに**2個の束の組合せ**が全部で何通りあるのかを考えてみる。ただし，結果的に同じ形の束になったとしても，それが作られる過程が異なれば異なる組合せとみなす。

$n=3$のときの2個の束の組合せの数は次のように求めることができる。$n=2$のときに2個の束をもつのは$(1,2)$の1通りだけである。$(1,\mathbf{2})$は2の左側に3を入れる2個のスペース（(③, 1, 2)，(1, ③, 2)）をもつ。また，1個の束をもつ組合せは$(2,1)$であり，$(\mathbf{2},1)$は2の右側に3を入れる2個のスペース（(2, 1, ③)，(2, ③, 1)）をもつ。よって，**$n=3$のときの2個の束の組合せは全部で$2+2=4$通りとなる。**

$n=4$のときの2個の束の組合せの数は次のように求めることができる。$n=3$のときに2個の束をもつのは$(1,3,2)$，$(2,1,3)$，$(2,3,1)$，$(3,1,2)$の4通りである。それぞれの場合について，3の左側に4を入れるスペースは，$(1,\mathbf{3},2)$は2個{(④, 1, 3, 2)，(1, ④, 3, 2)}，$(2,1,\mathbf{3})$は3個{(④, 2, 1, 3)，(2, ④, 1, 3)，(2, 1, ④, 3)}，$(2,\mathbf{3},1)$は2個{(④, 2, 3, 1)，(2, ④, 3, 1)}，$(\mathbf{3},1,2)$は1個{(④, 3, 1, 2)}もつから，合計8個のスペースをもつ。また，1個の束をもつ組合せは$(3,2,1)$であり，$(\mathbf{3},2,1)$は3の右側に4を入れる3個のスペース{(3, 2, 1, ④)，(3, 2, ④, 1)，(3, ④, 2, 1)}をもつ。よって，**$n=4$のときの2個の束の組合せは全部で$8+3=11$通りとなる。**

同様に，$n=5$のときに2個の束の組合せは全部で何通りあるかを調べてみる。$n=4$のときに2個の束をもつのは$(1,4,3,2)$，$(2,1,4,3)$，$(2,4,1,3)$，$(2,4,3,1)$，$(3,2,1,4)$，$(3,2,4,1)$，$(3,4,2,1)$，$(4,1,3,2)$，$(4,2,1,3)$，$(4,2,3,1)$，$(4,3,1,2)$の11通りである。それぞれの場合について，4の左側に5を入れるスペースは，$(1,\mathbf{4},3,2)$は2個，$(2,1,\mathbf{4},3)$は3個，$(2,\mathbf{4},1,3)$は2個，$(2,\mathbf{4},3,1)$は2個，$(3,2,1,\mathbf{4})$は4個，$(3,2,\mathbf{4},1)$は3個，$(3,\mathbf{4},2,1)$は2個，$(\mathbf{4},1,3,2)$は1個，$(\mathbf{4},2,1,3)$は1個，$(\mathbf{4},2,3,1)$は1個，$(\mathbf{4},3,1,2)$は1個もつから，合計22個

のスペースをもつ。また，1 個の束をもつ組合せは$(4, 3, 2, 1)$であり，$(\mathbf{4}, 3, 2, 1)$は 4 の右側に 5 を入れる 4 個のスペースをもつ。よって，**n＝5 のときの 2 個の束の組合せは全部で 22＋4＝26 通りとなる。**

この結果，n 枚のカードにおける 2 個の束の組合せの総数は，n－1 枚のカードにおいて束の数が不変となる組合せの数（2 個の束の組合せの数の 2 倍の数）と，束の数が 1 増加し 2 個の束になる組合せの数（2 個の束の組合せの数の n－2 倍の数）の和として得られることがわかる。

例えば，$n=3$ の 2 個の束の組合せは，$n-1=2$ の 2 個の束の組合せが 1 通りであるから，束の数が不変となる組合せは $1\times2=2$ 通りあり，これに 1 個の束をもつ場合から束の数が 1 増加し 2 個の束になる組合せの数 $n-1=2$ を加えることにより求められる。すなわち，$2+2=4$ 通りとなる。また，$n=4$ の 2 個の束の組合せは，$n-1=3$ の 2 個の束の組合せが 4 通りであるから，束の数が不変となる組合せは $4\times2=8$ 通りあり，これに 1 個の束をもつ場合から束の数が 1 増加し，2 個の束になる組合せの数 $n-1=3$ を加えることにより求められる。すなわち，$8+3=11$ 通りとなる。さらに，$n=5$ の 2 個の束の組合せは，$n-1=4$ の 2 個の束の組合せが 11 通りであるから，束の数が不変となる組合せは $11\times2=22$ 通りあり，これに 1 個の束をもつ場合から束の数が 1 増加し，2 個の束になる組合せの数 $n-1=4$ を加えることにより求められる。すなわち，$22+4=26$ 通りとなる。

束の数が 2 のとき，4 の左側にその組合せの数の 2 倍のスペースをもつ。一般に束の数が k のとき，4 の左側にその組合せの数の k 倍のスペースをもつ。

(1　2　3　4)…4	(2　1　3　4)…3	(3　1　2　4)…3	(4　1　2　3)…3
(1　2　4　3)…3	(2　1　④　3)…2	(3　1　4　2)…3	(④　1　3　2)…2
(1　3　2　4)…3	(2　3　1　4)…3	(3　2　1　④)…2	(④　2　1　3)…2
(1　3　4　2)…3	(2　3　4　1)…3	(3　2　④　1)…2	(④　2　3　1)…2
(1　4　2　3)…3	(2　④　1　3)…2	(3　4　1　2)…3	(④　3　1　2)…2
(1　④　3　2)…2	(2　④　3　1)…2	(3　④　2　1)…2	(4　3　2　1)…1

3.3　3 個の束の組合せ

いま，$n=4$，5 のときに **3 個の束の組合せ**が全部で何通りあるのかを考えてみる。

$n=4$ のときの 3 個の束の組合せの数は次のようにして求めることができる。$n=3$ のときに 3 個の束をもつのは$(1, 2, 3)$の 1 通りだけである。$(1, 2, \mathbf{3})$は 3 の左側に 4 を入れる 3 個のスペース｛(④, 1, 2, 3)，(1, ④, 2, 3)，(1, 2, ④, 3)｝をもつ。また，2 個の束をもつ組合せは$(1, 3, 2)$，$(2, 1, 3)$，$(2, 3, 1)$，$(3, 1, 2)$であり，$(1, \mathbf{3}, 2)$は 3 の右側に 4 を入れる 2 個のスペース｛(1, 3, 2, ④)，(1, 3, ④, 2)｝，$(2, 1, \mathbf{3})$は 3 の右側に 4 を入れる 1 個のスペース｛(2, 1, 3, ④)｝，$(2, \mathbf{3}, 1)$

は3の右側に4を入れる2個のスペース{(2, 3, 1, ④)，(2, 3, ④, 1)}，(**3**, 1, 2)は3の右側に4を入れる3個のスペース{(3, 1, 2, ④)，(3, 1, ④, 2)，(3, ④, 1, 2)}をもち，合計8個のスペースをもつ。よって，**n=4のときの3個の束の組合せは全部で3＋8＝11通りとなる。**

同様に，n=5のときに3個の束の組合せは全部で何通りあるかを調べてみる。n=4のときに3個の束をもつのは(1, 2, 4, 3)，(1, 3, 2, 4)，(1, 3, 4, 2)，(1, 4, 2, 3)，(2, 1, 3, 4)，(2, 3, 1, 4)，(2, 3, 4, 1)，(3, 1, 2, 4)，(3, 1, 4, 2)，(3, 4, 1, 2)，(4, 1, 2, 3)の11通りである。それぞれの場合について，4の左側に5を入れるスペースは，(1, 2, **4**, 3)は3個，(1, 3, 2, **4**)は4個，(1, 3, **4**, 2)は3個，(1, **4**, 2, 3)は2個，(2, 1, 3, **4**)は4個，(2, 3, 1, **4**)は4個，(2, 3, **4**, 1)は3個，(3, 1, 2, **4**)は4個，(3, 1, **4**, 2)は3個，(3, **4**, 1, 2)は2個，(**4**, 1, 2, 3)は1個もつから，合計33個のスペースをもつ。また，2個の束をもつ組合せのそれぞれの場合について，4の右側に5を入れるスペースは，(1, **4**, 3, 2)は3個，(2, 1, **4**, 3)は2個，(2, **4**, 1, 3)は3個，(2, **4**, 3, 1)は3個，(3, 2, 1, **4**)は1個，(3, 2, **4**, 1)は2個，(3, **4**, 2, 1)は3個，(**4**, 1, 3, 2)は4個，(**4**, 2, 1, 3)は4個，(**4**, 2, 3, 1)は4個，(**4**, 3, 1, 2)は4個もつから，合計33個のスペースをもつ。よって，**n=5のときの3個の束の組合せは全部で33＋33＝66通りとなる。**

この結果，n枚のカードにおける3個の束の組合せの総数は，n−1枚のカードにおいて束の数が不変となる組合せの数(3個の束の組合せの数の3倍の数)と，束の数が1増加し3個の束になる組合せの数(2個の束の組合せの数のn−2倍の数)の和として得られることがわかる。

例えば，n=4の3個の束の組合せは，n−1=3の3個の束の組合せが1通りであるから，束の数が不変となる組合せは1×3=3通りあり，これにn−1=3の2個の束の組合せが4通りであるから，束の数が1増加し3個の束になる組合せの数4×(4−2)=8を加えることにより求められる。すなわち，3+8=11通りとなる。さらに，n=5の3個の束の組合せは，n−1=4の3個の束の組合せの数が11通りであるから，束の数が不変となる組合せは11×3=33通りあり，これにn−1=4の2個の束の組合せが11通りであるから，束の数が1増加し3個の束になる組合せの数11×(5−2)=33を加えることにより求められる。すなわち，33+33=66通りとなる。

3.4　4個の束の組合せ

同様に考えると，n枚のカードにおける4個の束の組合せの総数は，n−1枚のカードにおいて束の数が不変となる組合せの数(4個の束の組合せの数の4倍の数)と，束の数が1増加し4個の束になる組合せの数(3個の束の組合せの数のn−3倍の数)の和として得られる。

3.5　n個の束の組合せ

一般的には，n枚のカードにおけるi個の束の組合せの総数は，n−1枚のカードにおいて

束の数が不変となる組合せの数（i 個の束の組合せの数の i 倍の数）と，束の数が 1 増加し i 個の束になる組合せの数（$i-1$ 個の束の組合せの数の $n-(i-1)$ 倍の数）の和として得られる。このことを定式化すると，以下のようになる。

一般に，n 枚のカードにおける i 個の束の組合せの数を $M_n(i)$ とすると，次の漸化式が成り立つ。

$$M_n(1)=M_n(n)=1 \quad (n\geq 1),$$
$$M_n(i)=iM_{n-1}(i)+\{n-(i-1)\}M_{n-1}(i-1) \quad (n\geq 3,\ i=2,\cdots,n-1).$$

したがって，n 枚のカードにおいて，i 個の束が作られる確率を $P_n(i)$ とすると，n 枚のカードの並べ方は $n!$ 通りあるから，$P_n(i)=M_n(i)/n!$ で与えられる。よって，以下の確率分布に関する漸化式が成り立つ。

$$P_n(1)=P_n(n)=\frac{1}{n!} \quad (n\geq 1),$$
$$P_n(i)=\frac{i}{n}P_{n-1}(i)+\frac{n-(i-1)}{n}P_{n-1}(i-1) \quad (n\geq 3,\ i=2,\cdots,n-1). \tag{1}$$

以下に，$M_n(i)$ および $P_n(i)$ の数表を与える。

$M_n(i)$ $(n=1, 2, \cdots, 12)$

$n\backslash i$	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	1	1										
3	1	4	1									
4	1	11	11	1								
5	1	26	66	26	1							
6	1	57	302	302	57	1						
7	1	120	1191	2416	1191	120	1					
8	1	247	4293	15619	15619	4293	247	1				
9	1	502	14608	88234	156190	88234	14608	502	1			
10	1	1013	47840	455192	1310354	1310354	455192	47840	1013	1		
11	1	2036	152637	2203488	9738114	15724248	9738114	2203488	152637	2036	1	
12	1	4083	478271	10187685	66318474	162512286	162512286	66318474	10187685	478271	4083	1

$P_n(i)$ $(n=1, 2, \cdots, 10)$

$n\backslash i$	1	2	3	4	5	6	7	8	9	10
1	1.0000000									
2	0.5000000	0.5000000								
3	0.1666667	0.6666667	0.1666667							
4	0.0416667	0.4583333	0.4583333	0.0416667						
5	0.0083333	0.2166667	0.5500000	0.2166667	0.0083333					
6	0.0013889	0.0791667	0.4194444	0.4194444	0.0791667	0.0013889				
7	0.0001984	0.0238095	0.2363095	0.4793651	0.2363095	0.0238095	0.0001984			
8	0.0000248	0.0061260	0.1064732	0.3873760	0.3873760	0.1064732	0.0061260	0.0000248		
9	0.0000028	0.0013834	0.0402557	0.2431493	0.4304178	0.2431493	0.0402557	0.0013834	0.0000028	
10	0.0000003	0.0002792	0.0131834	0.1254387	0.3610984	0.3610984	0.1254387	0.0131834	0.0002792	0.0000003

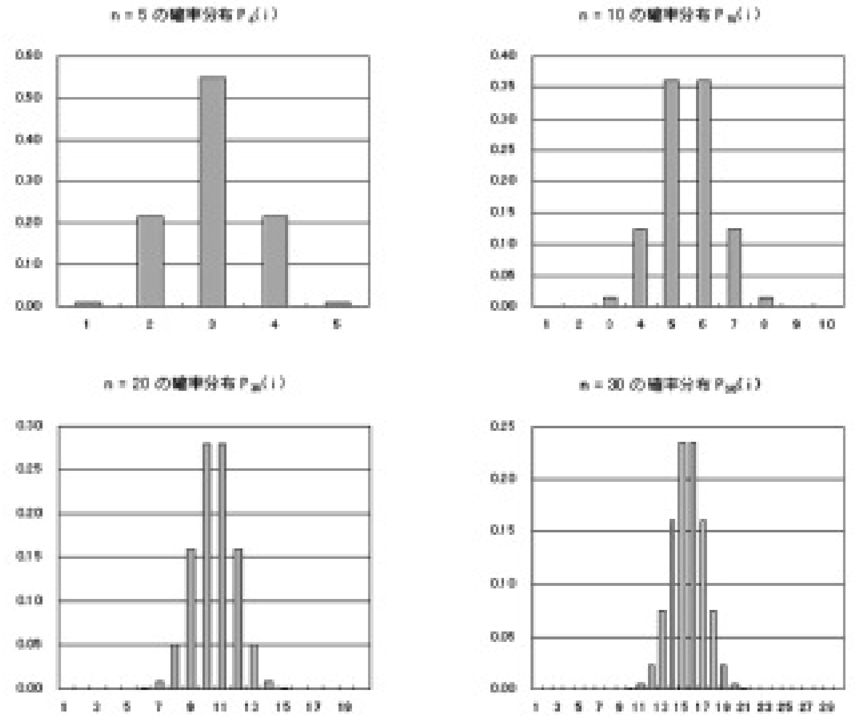


図 1：$P_n(i)$ $(n=5, 10, 20, 30)$の確率分布

4　確率分布のモーメント

本節では確率分布の期待値や分散をはじめとするモーメントを与え，その証明を示す。

定理 3.1　確率分布(1)式に従う確率変数 X の r 次モーメント $E_n(X^r)$は，以下の漸化式で与えられる。

$$E_1(X^r)=1,\quad E_n(X^r)=1+\sum_{k=1}^{r}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}E_{n-1}(X^k)\quad (n\geq 2). \tag{2}$$

証明　$n\geq 3$ のとき，X の r 次モーメント $E_n(X^r)$は(1)式より

$$E_n(X^r)=\sum_{i=1}^{n} i^r P_n(i)=\frac{1}{n!}+\frac{n^r}{n!}+\sum_{i=2}^{n-1}\frac{i^{r+1}}{n}P_{n-1}(i)+\sum_{i=2}^{n-1} i^r\frac{n-(i-1)}{n}P_{n-1}(i-1) \tag{3}$$

で与えられる。ここで，(3)式の右辺第 3 項と第 4 項はそれぞれ以下のように変形される。

$$\sum_{i=2}^{n-1}\frac{i^{r+1}}{n}P_{n-1}(i)=\frac{1}{n}\{E_{n-1}(X^{r+1})-P_{n-1}(1)\}=\frac{1}{n}E_{n-1}(X^{r+1})-\frac{1}{n!}\ , \tag{4}$$

$$\begin{aligned}
&\sum_{i=2}^{n-1} i^r \frac{n-i+1}{n} P_{n-1}(i-1)\\
&=\frac{n+1}{n}\sum_{i=2}^{n-1} i^r P_{n-1}(i-1)-\frac{1}{n}\sum_{i=2}^{n-1} i^{r+1} P_{n-1}(i-1)\\
&=\frac{n+1}{n}\sum_{i=2}^{n-1}\sum_{k=0}^{r}\binom{r}{k}(i-1)^k P_{n-1}(i-1)-\frac{1}{n}\sum_{i=2}^{n-1}\sum_{k=0}^{r+1}\binom{r+1}{k}(i-1)^k P_{n-1}(i-1)\\
&=\frac{n+1}{n}\sum_{k=0}^{r}\binom{r}{k}\{E_{n-1}(X^k)-(n-1)^{\mathrm{k}} P_{n-1}(n-1)\}\\
&\quad-\frac{1}{n}\sum_{k=0}^{r+1}\binom{r+1}{k}\{E_{n-1}(X^k)-(n-1)^k P_{n-1}(n-1)\}\\
&=\frac{n+1}{n}\sum_{k=0}^{r}\binom{r}{k}\{E_{n-1}(X^k)-(n-1)^k P_{n-1}(n-1)\}\\
&\quad-\frac{1}{n}\left[\sum_{k=0}^{r}\binom{r+1}{k}\{E_{n-1}(X^k)-(n-1)^k P_{n-1}(n-1)\}+E_{n-1}(X^{r+1})-(n-1)^{r+1}P_{n-1}(n-1)\right]\\
&=\sum_{k=0}^{r}\{E_{n-1}(X^k)-(n-1)^k P_{n-1}(n-1)\}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}\\
&\quad-\frac{1}{n}E_{n-1}(X^{r+1})+\frac{(n-1)^{r+1}}{n}P_{n-1}(n-1). \qquad (5)
\end{aligned}$$

よって，(4)式および(5)式を(3)式に代入することにより，

$$E_n(X^r)=\frac{n^r}{n!}+\sum_{k=0}^{r}\left\{E_{n-1}(X^k)-\frac{(n-1)^k}{(n-1)!}\right\}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}+\frac{(n-1)^{r+1}}{n!} \qquad (6)$$

を得る。ここで，(6)式において

$$\begin{aligned}
&\frac{n^r}{n!}-\sum_{k=0}^{r}\frac{(n-1)^k}{(n-1)!}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}+\frac{(n-1)^{r+1}}{n!}\\
&=\sum_{k=0}^{r}\binom{r}{k}\frac{(n-1)^k}{n!}-\sum_{k=0}^{r}\frac{(n-1)^k}{(n-1)!}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}+\binom{r+1}{r+1}\frac{(n-1)^{r+1}}{n!}\\
&=-\sum_{k=0}^{r}\binom{r}{k}\frac{n}{n!}(n-1)^k+\sum_{k=0}^{r+1}\binom{r+1}{k}\frac{(n-1)^k}{n!}\\
&=-\sum_{k=0}^{r}\binom{r}{k}\frac{(n-1)^{k+1}}{n!}-\sum_{k=0}^{r}\binom{r}{k}\frac{(n-1)^k}{n!}+\sum_{k=1}^{r}\left\{\binom{r}{k}+\binom{r}{k-1}\right\}\frac{(n-1)^k}{n!}+\frac{1}{n!}+\frac{(n-1)^{r+1}}{n!}\\
&=0.
\end{aligned}$$

ゆえに，$n\geq 3$ のとき r 次モーメント $E_n(X^r)$は，以下の漸化式で与えられる。

$$\begin{aligned}
E_n(X^r)&=\sum_{k=0}^{r}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}E_{n-1}(X^k)\\
&=1+\sum_{k=1}^{r}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}E_{n-1}(X^k). \qquad (7)
\end{aligned}$$

また，モーメントの定義より$E_1(X^r)=1^rP_1(1)=1$ であり，

$$E_2(X^r)=1^rP_2(1)+2^rP_2(2)=\frac{1}{2}(1+2^r) \qquad (8)$$

を得る。ここで，(8)式は(7)式で $n=2$ とおいた $E_2(X^r)$と一致する。したがって，(7)式は $n\geq 2$ のときにも成り立つ。 □

定理3.2　確率分布(1)に従う確率変数 X の期待値 $E_n(X)$ と分散 $V_n(X)$ は，それぞれ

$$E_n(X)=\frac{n+1}{2} \quad (n=1, 2, \cdots),$$

$$V_1(X)=0, \quad V_n(X)=\frac{n+1}{12} \quad (n=2, 3, \cdots)$$

で与えられる。

証明　$r=1$ のとき，(2)式は

$$E_1(X)=1, \quad E_n(X)=\frac{n-1}{n}E_{n-1}(X)+1 \tag{9}$$

と表される。$B_n(X)=nE_n(X)$ とおくと，(9)式は $B_1(X)=E_1(X)=1$,

$$B_n(X)=B_{n-1}(X)+n \tag{10}$$

となる。(10)式から $n\geq 2$ に対して，

$$B_n(X)=B_1(X)+\sum_{k=1}^{n-1}k=1+\frac{n(n-1)}{2}=\frac{n(n+1)}{2} \tag{11}$$

を得る。(11)式で $n=1$ のとき，1となり，$B_1(X)$ に一致する。よって，期待値 $E_n(X)$ の一般項は

$$E_n(X)=\frac{B_n(X)}{n}=\frac{n+1}{2} \quad (n=1, 2, \cdots)$$

で与えられる。

$r=2$ のとき，(2)式は

$$E_1(X^2)=1, \quad E_n(X^2)=\frac{n-2}{n}E_{n-1}(X^2)+\frac{2n-1}{n}E_{n-1}(X)+1 \tag{12}$$

と表される。よって，分散 $V_n(X)$ に関して $V_1(X)=E_1(X^2)-\{E_1(X)\}^2=0$ および

$$\begin{aligned}V_n(X)&=E_n(X^2)-\{E_n(X)\}^2=\frac{n-2}{n}E_{n-1}(X^2)-\frac{n^2}{4}+\frac{n}{2}+\frac{1}{4}\\&=\frac{n-2}{n}V_{n-1}(X)+\frac{1}{4}\end{aligned} \tag{13}$$

を得る。(13)式が

$$V_n(X)-(an+b)=\frac{n-2}{n}[V_{n-1}(X)-\{a(n-1)+b\}] \tag{14}$$

を満たすように定数 a, b の値を求めると，$a=b=1/12$ である。ここで，$C_n(X)=V_n(X)-\{(n+1)/12\}$ とおくと，(14)式は $C_n(X)=\{(n-2)/n\}C_{n-1}(X)$ となるから，$n\geq 2$ に対して $C_n(X)=0$ を得る。したがって，

$$V_n(X)=\frac{n+1}{12} \quad (n=2, 3, \cdots)$$

となる。 □

5　確率分布の近似

確率分布(1)式が正規分布の近似になっているので，本節はこれに関する議論をし，数値的にその程度を示す。

5.1　理論的背景

定理 4.1　確率分布(1)式に従う確率変数 X の 6 次までのキュムラント $\kappa_r^{(n)}$ $(r=1, 2, \cdots, 6)$は，

$$
\begin{aligned}
\kappa_1^{(n)} &= \frac{n+1}{2} && (n \geq 1),\\
\kappa_2^{(n)} &= \frac{n+1}{12} && (n \geq 2),\\
\kappa_3^{(n)} &= 0 && (n \geq 1),\\
\kappa_4^{(n)} &= -\frac{n+1}{120} && (n \geq 4),\\
\kappa_5^{(n)} &= 0 && (n \geq 1),\\
\kappa_6^{(n)} &= \frac{n+1}{252} && (n \geq 6)
\end{aligned}
$$

で与えられる。

証明　付録 A 参照。 □

X の確率分布は対称であるから，$r \geq 3$ の奇数次のキュムラントは 0 となる。また，1 次のキュムラントは$(n+1)(1+B_1)$，偶数次のキュムラントは$(n+1)B_r/r$ $(r \geq 2)$として表されることがわかる。ただし，B_r は次の関係式を満たすベルヌーイ数である。

$$\frac{t}{e^t-1} = \sum_{r=0}^{\infty} \frac{B_r}{r!}\, t^r.$$

$r=10$ までのベルヌーイ数 B_r は以下の通りである。

B_r $(r=1, 2, \cdots, 10)$

r	1	2	3	4	5	6	7	8	9	10
B_r	$-\frac{1}{2}$	$\frac{1}{6}$	0	$-\frac{1}{30}$	0	$\frac{1}{42}$	0	$-\frac{1}{30}$	0	$\frac{5}{66}$

したがって，確率分布(1)式は $n \geq r$ の r 次のキュムラント $\kappa_r^{(n)}$ が$(0, 1)$上の一様分布に従う $n+1$ 個の互いに独立で同一な確率変数の和のキュムラントと一致することがわかる。

定理 4.2　確率変数 T_n が確率分布(1)式に従うとき，

$$X_n = \frac{\sqrt{N}\,(T_n - N/2)}{N/\sqrt{12}}$$

とおくと，$n \geq 6$ の X_n の分布関数に対して，次の漸近展開式を得る。

$$\Pr[X_n < x] = \Phi(x) - \phi(x)\left\{\frac{1}{N}a_1(x) + \frac{1}{N^2}a_2(x) + O(N^{-3})\right\}. \tag{15}$$

ここで，$N = n+1$, $\Phi(x)$ と $\phi(x)$ はそれぞれ標準正規分布関数と標準正規密度関数，係数 $a_1(x)$, $a_2(x)$ はそれぞれ以下のように与えられる。

$$a_1(x) = -\frac{1}{20}H_3(x),$$

$$a_2(x) = \frac{1}{105}H_5(x) + \frac{1}{800}H_7(x).$$

ただし，$H_j(x)(j=1, 2, \cdots)$ は j 次のエルミート多項式である。

証明　$N = n+1$ として

$$X_n = \frac{\sqrt{N}\,(T_n - \mu)}{\sigma}$$

とおく。ただし，$\mu = N/2$，$\sigma = N/\sqrt{12}$ とする。このとき，X_n の特性関数は次の式で与えられる。

$$\psi_{X_n}(t) = E(e^{itX_n}) = \exp\left(-\frac{\sqrt{N}\,\mu}{\sigma}it\right)E\left[\exp\left(\frac{\sqrt{N}\,T_n}{\sigma}it\right)\right]. \tag{16}$$

T_n の特性関数

$$\psi_{T_n}(t) = E(e^{itT_n}) = \exp\left[(it)\kappa_1^{(n)} + \frac{(it)^2}{2}\kappa_2^{(n)} + \frac{(it)^3}{3!}\kappa_3^{(n)} + \frac{(it)^4}{4!}\kappa_4^{(n)} + \cdots\right]$$

とおくと，(16)式は

$$\begin{aligned}\psi_{X_n}(t) &= \exp\left(-\frac{\sqrt{N}\,\mu}{\sigma}it\right)\psi_{T_n}\left(\frac{\sqrt{N}}{\sigma}t\right)\\ &= \exp\left(-\frac{\sqrt{N}\,\mu}{\sigma}it\right)\exp\left[\left(\frac{\sqrt{N}}{\sigma}it\right)\kappa_1^{(n)} + \frac{1}{2}\left(\frac{\sqrt{N}}{\sigma}it\right)^2\kappa_2^{(n)}\right.\\ &\qquad\left. + \frac{1}{3!}\left(\frac{\sqrt{N}}{\sigma}it\right)^3\kappa_3^{(n)} + \frac{1}{4!}\left(\frac{\sqrt{N}}{\sigma}it\right)^4\kappa_4^{(n)} + \cdots\right]\end{aligned} \tag{17}$$

と表される。

　ここで，(17)式に $\kappa_1^{(n)} = \mu = N/2$，$\kappa_2^{(n)} = \sigma^2/N = N/12$，$\kappa_3^{(n)} = 0$，$\kappa_4^{(n)} = -N/120$, $\kappa_5^{(n)} = 0$, $\kappa_6^{(n)} = N/252$ を代入すると，

$$\psi_{X_n}(t) = \exp\left(-\frac{t^2}{2}\right)\left[1 - \frac{1}{N}\left\{\frac{1}{20}(it)^4\right\} + \frac{1}{N^2}\left\{\frac{1}{105}(it)^6 + \frac{1}{800}(it)^8\right\} + O(N^{-3})\right] \tag{18}$$

となる。いま，$\phi(x)$ を標準正規密度関数とし，

$$\frac{1}{2\pi}\int(it)^r e^{-t^2/2}e^{-itx}dx=H_r(x)\phi(x)$$

を用いて(18)式を反転する。ただし，$H_r(x)$は

$$\left(-\frac{d}{dx}\right)^r\phi(x)=H_r(x)\phi(x)$$

で与えられるエルミート多項式である。その結果，次の式を得る。

$$\Pr[X_n<x]=\Phi(x)-\phi(x)\left[-\frac{1}{N}\left\{\frac{1}{20}H_3(x)\right\}+\frac{1}{N^2}\left\{\frac{1}{105}H_5(x)+\frac{1}{800}H_7(x)\right\}+O(N^{-3})\right]. \quad (19)$$

□

系 4.3　確率変数 T_n が確率分布(1)式に従うとき，$n\geq 6$ の X_n の密度関数 $f(x)$に対して，次の漸近展開式を得る。

$$f(x)=\phi(x)\left[1-\frac{1}{N}\left\{\frac{1}{20}H_4(x)\right\}+\frac{1}{N^2}\left\{\frac{1}{105}H_6(x)+\frac{1}{800}H_8(x)\right\}+O(N^{-3})\right]. \quad (20)$$

証明　(19)式の両辺を x に関して微分すればよい。　□

5.2　数値近似

図 2 は確率分布(1)式の値を $n=6, 10, 20$ の場合について，次式の第 1 項(正規近似)，$1/N$ の項，$1/N^2$ の項までの近似式 R, Z1, Z2 を用いて計算し，それらと真の値の差を表したものである。

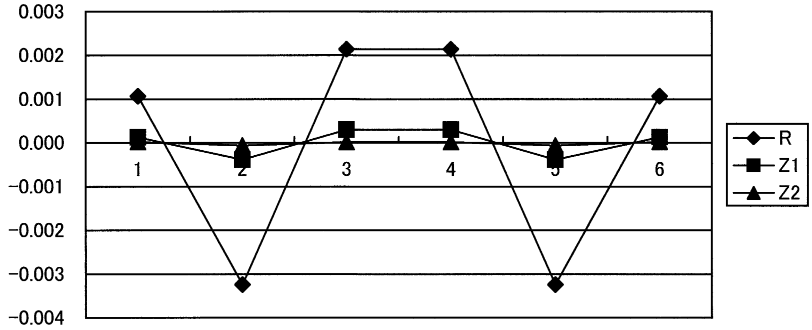
$$P(T_n=i)=\frac{1}{\sqrt{2\pi N/12}}e^{-\frac{(i-N/2)^2}{N/6}}\left[1-\frac{1}{N}\left\{\frac{1}{20}H_4(x)\right\}\right.$$
$$\left.+\frac{1}{N^2}\left\{\frac{1}{105}H_6(x)+\frac{1}{800}H_8(x)\right\}+O(N^{-3})\right]. \quad (21)$$

これを見ると，近似式 R でも誤差が±0.003 程度の範囲内となっており，十分良い近似値を与えているが，近似式 Z1, Z2 により確率分布のとり得るすべての値について近似精度が改善されていることがわかる。

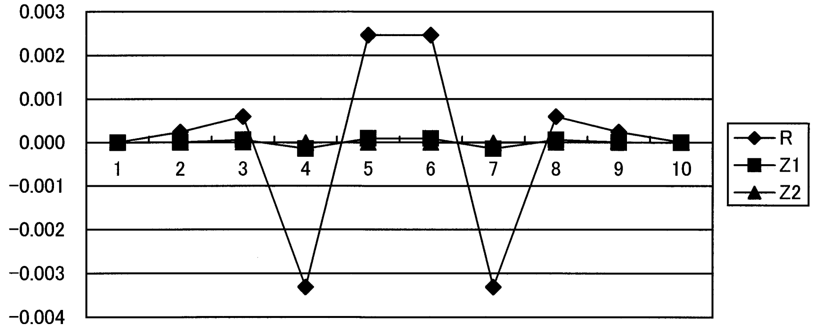
参 考 文 献

[1] 竹内　啓 (1975). 確率分布の近似―シリーズ 新しい応用の数学 10，教育出版.

$n=6$ の確率分布



$n=10$ の確率分布



$n=20$ の確率分布

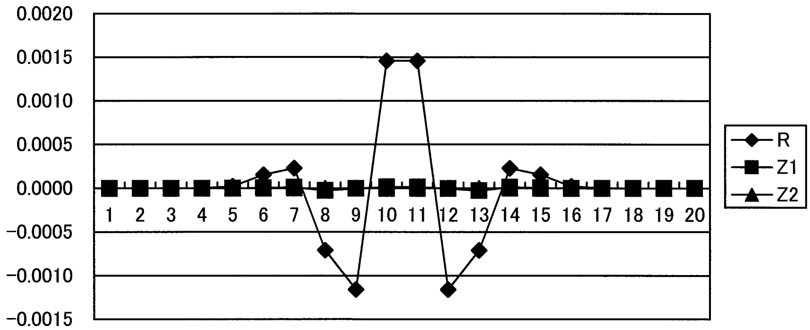


図 2：n＝6, 10, 20 の確率分布の値を(21)式の第 1 項(正規近似)，$1/N$ の項，$1/N^2$ の項まで用いて計算したときの誤差＝近似値－真の値(それぞれ，R, Z1, Z2 で表した)

付　録

A　定理 4.1 の証明

A.1　キュムラントとモーメントの関係

キュムラント $\kappa_r^{(n)}$ とモーメント $E_n(X^r)$の関係

$$\exp\left\{\sum_{r=1}^{\infty}\frac{t^r}{r!}\kappa_r^{(n)}\right\}=\sum_{r=0}^{\infty}\frac{t^r}{r!}E_n(X^r)$$

から，キュムラントとモーメントの $r=6$ までの関係は以下のように与えられる。ただし，$\kappa_r^{(n)}=\kappa_r, E_n(X^r)=\mu_r'$ とする。

$$\begin{aligned}
\mu_1'&=\kappa_1,\\
\mu_2'&=\kappa_2+\kappa_1^2,\\
\mu_3'&=\kappa_3+3\kappa_2\kappa_1+\kappa_1^3,\\
\mu_4'&=\kappa_4+4\kappa_3\kappa_1+3\kappa_2^2+6\kappa_2\kappa_1^2+\kappa_1^4,\\
\mu_5'&=\kappa_5+5\kappa_4\kappa_1+10\kappa_3\kappa_2+10\kappa_3\kappa_1^2+15\kappa_2^2\kappa_1+10\kappa_2\kappa_1^3+\kappa_1^5,\\
\mu_6'&=\kappa_6+6\kappa_5\kappa_1+15\kappa_4\kappa_2+15\kappa_4\kappa_1^2+10\kappa_3^2+60\kappa_3\kappa_2\kappa_1+20\kappa_3\kappa_1^3+15\kappa_2^3+45\kappa_2^2\kappa_1^2+15\kappa_2\kappa_1^4+\kappa_1^6.
\end{aligned}$$

A.2　1 次，2 次のキュムラント

定理 3.2 より明らか。

A.3　3 次のキュムラント

X の確率分布は対称であるから，3 次のキュムラントは 0 となる。実際，$r=3$ のとき，(2) 式は

$$\begin{aligned}
&E_1(X^3)=1,\\
&E_n(X^3)=\frac{n-3}{n}E_{n-1}(X^3)+\frac{3n-3}{n}E_{n-1}(X^2)+\frac{3n-1}{n}E_{n-1}(X)+1
\end{aligned}\tag{22}$$

と表される。ここで，定理 3.2 の結果から

$$E_n(X^2)=\frac{1}{4}n^2+\frac{7}{12}n+\frac{1}{3}\quad(n\geq 2)$$

を用いると，(22)式は $n\geq 3$ に対して

$$E_n(X^3)=\frac{n-3}{n}E_{n-1}(X^3)+\frac{3}{4}n^2+n+\frac{1}{4}\tag{23}$$

となる。よって，$\kappa_3^{(n)}$ に関して

$$\begin{aligned}\kappa_3^{(n)}&=E_n(X^3)-3\kappa_2^{(n)}\kappa_1^{(n)}-\{\kappa_1^{(n)}\}^3\\&=\frac{n-3}{n}\kappa_3^{(n-1)}\end{aligned}\tag{24}$$

を得るから，(24)式は $n\geq 3$ に対して $\kappa_3^{(n)}=0$ となる。また，$n=1$ のとき，$E_1(X^3)=1$，$\kappa_2^{(1)}=0$，$\kappa_1^{(1)}=1$ より，

$$\kappa_3^{(1)}=E_1(X^3)-3\kappa_2^{(1)}\kappa_1^{(1)}-\{\kappa_1^{(1)}\}^3=0.$$

$n=2$ のとき，$E_2(X^3)=9/2$，$\kappa_2^{(2)}=1/4$，$\kappa_1^{(2)}=3/2$ より，

$$\kappa_3^{(2)}=E_2(X^3)-3\kappa_2^{(2)}\kappa_1^{(2)}-\{\kappa_1^{(2)}\}^3=0.$$

したがって，3 次のキュムラントは

$$\kappa_3^{(n)}=0\quad(n\geq 1)$$

となる。

A.4　4 次のキュムラント

$r=4$ のとき，(2)式は

$$\begin{aligned}&E_1(X^4)=1,\\&E_n(X^4)=\frac{n-4}{n}E_{n-1}(X^4)+\frac{4n-6}{n}E_{n-1}(X^3)+\frac{6n-4}{n}E_{n-1}(X^2)+\frac{4n-1}{n}E_{n-1}(X)+1\end{aligned}\tag{25}$$

と表される。ここで，3 次のキュムラントの結果から

$$E_n(X^3)=\frac{1}{8}n^3+\frac{1}{2}n^2+\frac{5}{8}n+\frac{1}{4}\quad(n\geq 2)$$

を用いると，(25)式は $n\geq 3$ に対して

$$E_n(X^4)=\frac{n-4}{n}E_{n-1}(X^4)+\frac{1}{2}n^3+\frac{5}{4}n^2+\frac{3}{4}n+\frac{1}{6}\tag{26}$$

となる。よって，$\kappa_4^{(n)}$ に関して

$$\begin{aligned}\kappa_4^{(n)}&=E_n(X^4)-3\{\kappa_2^{(n)}\}^2-6\kappa_2^{(n)}\{\kappa_1^{(n)}\}^2-\{\kappa_1^{(n)}\}^4\\&=\frac{n-4}{n}\kappa_4^{(n-1)}-\frac{1}{24}\end{aligned}\tag{27}$$

を得る。(27)式が

$$\kappa_4^{(n)}-(an+b)=\frac{n-4}{n}[\kappa_4^{(n-1)}-\{a(n-1)+b\}]\tag{28}$$

を満たすように定数 a, b の値を求めると，$a=b=-1/120$ である。ここで，$C_n(X)=\kappa_4^{(n)}+\{(n+1)/120\}$ とおくと，(28)式は $C_n(X)=\{(n-4)/n\}C_{n-1}(X)$となるから，$n\geq 4$ に対して $C_n(X)=0$ を得る。したがって，4 次のキュムラントは

$$\kappa_4^{(n)}=-\frac{n+1}{120}\quad (n\geq 4)$$

となる。

A.5　5次のキュムラント

X の確率分布は対称であるから，5次のキュムラントは0となる。実際，$r=5$ のとき，(2)式は

$$\begin{aligned}
&E_1(X^5)=1,\\
&E_n(X^5)=\frac{n-5}{n}E_{n-1}(X^5)+\frac{5n-10}{n}E_{n-1}(X^4)+\frac{10n-10}{n}E_{n-1}(X^3)\\
&\qquad +\frac{10n-5}{n}E_{n-1}(X^2)+\frac{5n-1}{n}E_{n-1}(X)+1
\end{aligned}\tag{29}$$

と表される。ここで，4次のキュムラントの結果から

$$E_n(X^4)=\frac{1}{16}n^4+\frac{3}{8}n^3+\frac{37}{48}n^2+\frac{79}{120}n+\frac{1}{5}\quad (n\geq 4)$$

を用いると，(29)式は $n\geq 5$ に対して

$$E_n(X^5)=\frac{n-5}{n}E_{n-1}(X^5)+\frac{5}{16}n^4+\frac{5}{4}n^3+\frac{65}{48}n^2+\frac{7}{12}n+\frac{1}{6}\tag{30}$$

となる。よって，$\kappa_5^{(n)}$ に関して

$$\begin{aligned}
\kappa_5^{(n)}&=E_n(X^5)-5\kappa_4^{(n)}\kappa_1^{(n)}-15\{\kappa_2^{(n)}\}^2\kappa_1^{(n)}-10\kappa_2^{(n)}\{\kappa_1^{(n)}\}^3-\{\kappa_1^{(n)}\}^5\\
&=\frac{n-5}{n}\kappa_5^{(n-1)}
\end{aligned}\tag{31}$$

を得るから，(31)式は $n\geq 5$ に対して $\kappa_5^{(n)}=0$ となる。また，$n=1$ のとき，$E_1(X^5)=1$，$\kappa_4^{(1)}=0$ より，

$$\kappa_5^{(1)}=E_1(X^5)-5\kappa_4^{(1)}\kappa_1^{(1)}-15\{\kappa_2^{(1)}\}^2\kappa_1^{(1)}-10\kappa_2^{(1)}\{\kappa_1^{(1)}\}^3-\{\kappa_1^{(1)}\}^5=0.$$

$n=2$ のとき，$E_2(X^5)=33/2$，$\kappa_4^{(2)}=-1/8$ より，

$$\kappa_5^{(2)}=E_2(X^5)-5\kappa_4^{(2)}\kappa_1^{(2)}-15\{\kappa_2^{(2)}\}^2\kappa_1^{(2)}-10\kappa_2^{(2)}\{\kappa_1^{(2)}\}^3-\{\kappa_1^{(2)}\}^5=0.$$

$n=3$ のとき，$E_3(X^5)=62$，$E_3(X^4)=73/3$，$\kappa_4^{(3)}=0$ より，

$$\kappa_5^{(3)}=E_3(X^5)-5\kappa_4^{(3)}\kappa_1^{(3)}-15\{\kappa_2^{(3)}\}^2\kappa_1^{(3)}-10\kappa_2^{(3)}\{\kappa_1^{(3)}\}^3-\{\kappa_1^{(3)}\}^5=0.$$

$n=4$ のとき，$E_4(X^5)=675/4$，$E_4(X^4)=331/6$，$\kappa_4^{(4)}=-1/24$ より，

$$\kappa_5^{(4)}=E_4(X^5)-5\kappa_4^{(4)}\kappa_1^{(4)}-15\{\kappa_2^{(4)}\}^2\kappa_1^{(4)}-10\kappa_2^{(4)}\{\kappa_1^{(4)}\}^3-\{\kappa_1^{(4)}\}^5=0.$$

したがって，5次のキュムラントは

$$\kappa_5^{(n)}=0\quad (n\geq 1)$$

となる。

A.6　6次のキュムラント

$r=6$ のとき，(2)式は

$$
\begin{aligned}
&E_1(X^6)=1,\\
&E_n(X^6)=\frac{n-6}{n}E_{n-1}(X^6)+\frac{6n-15}{n}E_{n-1}(X^5)+\frac{15n-20}{n}E_{n-1}(X^4)\\
&\qquad+\frac{20n-15}{n}E_{n-1}(X^3)+\frac{15n-6}{n}E_{n-1}(X^2)+\frac{6n-1}{n}E_{n-1}(X)+1 \qquad (32)
\end{aligned}
$$

と表される。ここで，5次のキュムラントの結果から

$$E_n(X^5)=\frac{1}{32}n^5+\frac{25}{96}n^4+\frac{25}{32}n^3+\frac{103}{96}n^2+\frac{11}{16}n+\frac{1}{6} \quad (n\geq 4)$$

を用いると，(32)式は $n\geq 5$ に対して

$$E_n(X^6)=\frac{n-6}{n}E_{n-1}(X^6)+\frac{3}{16}n^5+\frac{35}{32}n^4+\frac{15}{8}n^3+\frac{41}{32}n^2+\frac{31}{48}n+\frac{1}{6} \qquad (33)$$

となる。よって，$\kappa_6^{(n)}$ に関して

$$
\begin{aligned}
\kappa_6^{(n)}&=E_n(X^6)-15\kappa_4^{(n)}\kappa_2^{(n)}-15\kappa_4^{(n)}\{\kappa_1^{(n)}\}^2-15\{\kappa_2^{(n)}\}^3\\
&\quad-45\{\kappa_2^{(n)}\}^2\{\kappa_1^{(n)}\}^2-15\kappa_2^{(n)}\{\kappa_1^{(n)}\}^4-\{\kappa_1^{(n)}\}^6\\
&=\frac{n-6}{n}\kappa_6^{(n-1)}+\frac{1}{36} \qquad (34)
\end{aligned}
$$

を得る。(34)式が

$$\kappa_6^{(n)}-(an+b)=\frac{n-6}{n}[\kappa_6^{(n-1)}-\{a(n-1)+b\}] \qquad (35)$$

を満たすように定数 a，b の値を求めると，$a=b=1/252$ である。ここで，$C_n(X)=\kappa_6^{(n)}-\{(n+1)/252\}$ とおくと，(35)式は $C_n(X)=\{(n-6)/n\}C_{n-1}(X)$ となるから，$n\geq 6$ に対して $C_n(X)=0$ を得る。したがって，6次のキュムラントは

$$\kappa_6^{(n)}=\frac{n+1}{252} \quad (n\geq 6)$$

となる。

（つちや　たかひろ　統計科学専攻）
（なかむら　ながとも　統計科学，計量分析学専攻）

（2007年9月5日受理）