2008 年 05 月 24 日

配列内のランダムな配置の選択の評価

新潟工科大学 情報電子工学科 竹野茂治

1 はじめに

学生に

> 「N 個の要素のある配列のうち、乱数を使ってそのうちの m 個 $(1 \leq m \leq N)$ をランダムに 1 に、残りを 0 にする」

というプログラムを書いてもらったところ、どうも以下のように考えることが多いようであった (これを方法 A と呼ぶ):

1. 最初に N 個全部を 0 にする
2. $k = 0$ とし、$k = m$ になるまで以下を繰り返す:

   N 個の一つをランダムに選び、そこが 0 ならば 1 として k を一つ増やすが、そこが 1 ならば何もしない

しかしこの方法 A の場合、既に 1 であるところにぶつかるとまた乱数を取ってやり直すので、何回で終わるという保証はないし、極端な話、乱数があまりよくない乱数である場合には、かなりの回数がかかってしまう可能性もある。

よって以下のように考えるのがまだいいのではないかと思う (これを方法 B と呼ぶ):

1. 最初に N 個全部を 0 にする
2. $k = 0$ とし、$k = m$ になるまで以下を繰り返す:

   現在 0 の箇所の $(N - k)$ 個から一つをランダムに選び、そこを 1 として k を一つ増やす

この方法 B なら、もちろん丁度 m 回で終わるし、乱数も m 回生成するだけで済む。

では、方法 A の場合は、終了するまでだいたい何回位かかるのであろうか。本稿ではそれを考察してみることにする。

2 問題設定と簡単な場合の考察

問題を以下のように設定する:

> 1 から N までの N 個の数字から、ランダムに一つを選ぶことを繰り返す。一回の選択でそのいずれが出る確率も等しく $1/N$ であるとする。選んだ数の集合の異なる要素の数が m $(1 \leq m \leq N)$ になるまでそれを繰り返すこととする。それが終わるまでに繰り返し選択した回数の期待値 (平均値) $\mu(N, m)$ を求めよ。

まず、簡単のために、$N = 10,\ m = 3$ として考えてみることにする。この場合は、最低 3 回の選択が必要になるが、3 回で終わるのは、最初の一つ目の数はどれでもよく、二つ目は一つ目と違う数であればよく、三つ目の数は前の 2 つと違うものであればよいので、3 回で終わる確率 p_3 は、

$$p_3 = 1 \times \frac{9}{10} \times \frac{8}{10} = \frac{72}{100}$$

である。

4 回で終わるのは、2 回目に一つ目の数と同じものが出てしまうか、または 3 回目に一つ目か二つ目の数と同じものが出る場合なので、その確率 p_4 は

$$p_4 = 1 \times \frac{1}{10} \times \frac{9}{10} \times \frac{8}{10} + 1 \times \frac{9}{10} \times \frac{2}{10} \times \frac{8}{10} = \frac{72}{100} \times \frac{3}{10}$$

となる。

同様に、5 回の場合は、(2 回目、3 回目) が前と同じものか、(2 回目、4 回目) が、または (3 回目、4 回目) が前にでたものと同じものか、のいずれかの場合であり、よって、

$$p_5 = \frac{72}{100} \times \left(\frac{1}{10} \times \frac{1}{10} + \frac{1}{10} \times \frac{2}{10} + \frac{2}{10} \times \frac{2}{10}\right)$$

のようになる。これを繰り返すと、$(k+3)$ 回 $(k \geq 0)$ となる確率 p_{k+3} は、

$$\begin{aligned} p_{k+3} &= \frac{72}{100}\left\{\left(\frac{1}{10}\right)^k + \left(\frac{1}{10}\right)^{k-1}\frac{2}{10} + \left(\frac{1}{10}\right)^{k-2}\left(\frac{2}{10}\right)^2 + \cdots + \left(\frac{2}{10}\right)^k\right\} \\ &= \frac{72}{100}\sum_{j=0}^{k}\left(\frac{1}{10}\right)^{k-j}\left(\frac{2}{10}\right)^j \end{aligned}$$

となることがわかる。これは公比 2 の等比級数なので、

$$p_{k+3} = \frac{72}{100} \times \left(\frac{1}{10}\right)^k \frac{2^{k+1}-1}{2-1} = \frac{72}{100} \times \frac{2^{k+1}-1}{10^k}$$

となる。よって、回数の期待値 $\mu(10,3)$ は、

$$\begin{aligned}
\mu(10,3) &= \sum_{k=0}^{\infty}(k+3)p_{k+3} = \frac{72}{100}\sum_{k=0}^{\infty}(k+3)\frac{2^{k+1}-1}{10^k} \\
&= \frac{72}{100}\sum_{k=0}^{\infty}\left\{\frac{4}{10}k\left(\frac{2}{10}\right)^{k-1} - \frac{1}{10}k\left(\frac{1}{10}\right)^{k-1} + 6k\left(\frac{2}{10}\right)^k + 3k\left(\frac{1}{10}\right)^k\right\}
\end{aligned}$$

となる。ここで、$|x| < 1$ に対して、

$$\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \tag{1}$$

であるから (無限等比級数)、この両辺を微分すれば

$$\sum_{k=1}^{\infty} kx^{k-1} = 1 + 2x + 3x^2 + 4x^3 + \cdots = \frac{1}{(1-x)^2} \tag{2}$$

が成り立つことに注意する。

よって、この (1), (2) より

$$\begin{aligned}
\sum_{k=0}^{\infty} k\left(\frac{2}{10}\right)^{k-1} &= \sum_{k=1}^{\infty} k\left(\frac{2}{10}\right)^{k-1} = \frac{1}{(8/10)^2} = \frac{100}{64}, \\
\sum_{k=0}^{\infty} k\left(\frac{1}{10}\right)^{k-1} &= \sum_{k=1}^{\infty} k\left(\frac{1}{10}\right)^{k-1} = \frac{1}{(9/10)^2} = \frac{100}{81}, \\
\sum_{k=0}^{\infty} k\left(\frac{2}{10}\right)^{k} &= \frac{1}{8/10} = \frac{10}{8}, \quad \sum_{k=0}^{\infty} k\left(\frac{1}{10}\right)^{k} = \frac{1}{9/10} = \frac{10}{9}
\end{aligned}$$

なので、

$$\begin{aligned}
\mu(10,3) &= \frac{72}{100}\left(\frac{4}{10} \times \frac{100}{64} - \frac{1}{10} \times \frac{100}{81} + 6 \times \frac{10}{8} - 3 \times \frac{10}{9}\right) \\
&= \frac{9}{20} - \frac{8}{90} + \frac{54}{10} - \frac{24}{10} = \frac{81-16}{180} + 3 = 3 + \frac{13}{36}
\end{aligned}$$

となる。

3 一般化について

2 節では $N = 10$, $m = 3$ の場合について考察したが、しかしこれを一般化するのは容易ではない。例えば $N = 10$, $m = 4$ の場合を考えてみる。

この場合、例えば 7 回で終わる (3 回余計にかかる) のは、3 回分前に出ているものと同じものが 6 回目までに出る場合で、それぞれの確率は、それまでに出ている数の個数に応じて 1/10, 2/10, 3/10 となる。よって、p_7 は、

$$p_7 = \frac{9}{10}\frac{8}{10}\frac{7}{10}\sum_{1\le i_1\le i_2\le i_3\le 3}\frac{i_1}{10}\frac{i_2}{10}\frac{i_3}{10}$$

となる。この場合は、$1 \le i_1 \le i_2 \le i_3 \le 3$ となるすべての (i_1, i_2, i_3) の組を上げて数えてみれば求められなくはないが、一般には、

$$p_{k+4} = \frac{9}{10}\frac{8}{10}\frac{7}{10}\sum_{1\le i_1\le\ldots\le i_k\le 3}\frac{i_1}{10}\cdots\frac{i_k}{10}$$

であり、この最後の和を求めるのは容易ではない。一応、

$$S(k,r) = \sum_{1\le i_1\le\ldots\le i_k\le r} i_1\cdots i_k$$

とすると、

$$\begin{aligned} S(k,r) &= \sum_{i_k=1}^{r} i_k \sum_{1\le i_1\le\ldots\le i_{k-1}\le i_k} i_1\cdots i_{k-1} = \sum_{i=1}^{r} iS(k-1,i), \\ S(1,r) &= \sum_{1\le i\le r} i = \frac{1}{2}r(r+1), \quad S(k,1) = 1 \end{aligned}$$

などの性質を用いれば、小さい r に対しては $S(k,r)$ を順次決定できなくはないが、かなり大変である (詳しくは 6 節を参照)。

よって、一般の場合を考える場合は、別な方法をとる方がよい。

4 次の要素が得られるまでの回数の期待値

この節では、2 節、3 節とは異なる方法で、期待値 $\mu(N,m)$ を計算してみることにする。

この問題では、一つ目はどれを選んでもいいので一つ目は必ず 1 回で決まる。この先、次の二つ目 (2 種類目) の数字が出るまでに平均何回かかるかを考えてみる。

この場合は、一つ目の同じものが出続けている間は選択し続けるが、そうでないものが出ればそこで終わりである。つまり、

$$\begin{aligned}
1\ \text{回で済む確率} &= \frac{N-1}{N}, \\
2\ \text{回で済む確率} &= \frac{1}{N} \times \frac{N-1}{N}, \\
3\ \text{回で済む確率} &= \left(\frac{1}{N}\right)^2 \times \frac{N-1}{N}
\end{aligned}$$

のようになる。よって、二つ目の数字が出るまでの回数の期待値は、

$$\mu_\beta = \sum_{k=1}^{\infty} k \left(\frac{1}{N}\right)^{k-1} \frac{N-1}{N} = \frac{N-1}{N} \frac{1}{(1-1/N)^2} = \frac{N}{N-1}$$

となる。

$(j-1)$ 個の異なる数字が取れたところから始めると、次の j 個目の異なる数字が取れるまでには、

$$\begin{aligned}
1\ \text{回で済む確率} &= \frac{N-j+1}{N}, \\
2\ \text{回で済む確率} &= \frac{j-1}{N} \times \frac{N-j+1}{N}, \\
3\ \text{回で済む確率} &= \left(\frac{j-1}{N}\right)^2 \times \frac{N-j+1}{N}
\end{aligned}$$

のようになるので、期待値は

$$\begin{aligned}
\mu_j &= \sum_{k=1}^{\infty} k \left(\frac{j-1}{N}\right)^{k-1} \frac{N-j+1}{N} = \frac{N-j+1}{N} \frac{1}{\{1-(j-1)/N\}^2} \\
&= \frac{N}{N-j+1}
\end{aligned}$$

となる。$\mu(N,m)$ は、この μ_j を $j=1$ から $j=m$ まで累積したものになるので、

$$\mu(N,m) = \sum_{j=1}^{m} \frac{N}{N-j+1} \tag{3}$$

となる。例えば、$N=10,\ m=3$ の場合は、

$$\mu(10,3)=\frac{10}{10}+\frac{10}{9}+\frac{10}{8}=3+\frac{1}{9}+\frac{1}{4}=3+\frac{13}{36}$$

となって、2 節の結果と確かに一致する。

5 期待値の大きさの評価

この節では (3) を用いて、期待値 $\mu(N,m)$ の大きさを考えてみる。

例えば $N=10$ の場合、各 m に対する期待値 $\mu(10,m)$ の値は、表 1 のようになる。

m	1	2	3	4	5	6	7	8	9	10
$\mu(10,m)$	1.00	2.11	3.36	4.79	6.46	8.46	10.96	14.29	19.29	29.29

表 1: $N=10$ の場合の期待値

この $N=10$ の場合は、$m=8$ でも回数の期待値は m の 2 倍にもならず、それほど大きいわけではない。N が大きい場合でも、m があまり大きくなければ m と $\mu(N,m)$ とはそれほど離れない。

実際、$N\rightarrow\infty$ のときに m も N との比を固定して大きくする、例えば $m=[rN]$ $(0<r<1)$ のようにすると $m/N\rightarrow r$ であり、(3) より、

$$\begin{aligned}\frac{\mu(N,m)}{N} &= \sum_{k=1}^{m}\frac{1}{N-k+1}=\sum_{k=1}^{m}\frac{1}{N}\,\frac{1}{1-(k-1)/N}\\ &\rightarrow \int_0^r\frac{dx}{1-x}=\log\frac{1}{1-r}\quad(0<r<1)\end{aligned}$$

となることがわかる。つまり、大きい N に対しては、$r=m/N$ に対して

$$\frac{\mu(N,m)}{N}\approx\log\frac{1}{1-r}$$

であるので、

$$\frac{\mu(N,m)}{m}\approx\frac{1}{r}\log\frac{1}{1-r}$$

となるが、この最後の関数 $f(r) = (1/r)\log\{1/(1-r)\}$ は $0 \le r < 1$ で増加関数で、$f(+0) = 1$, $f(0.8) = 2.01$ 位なので、$m < 0.8N$ なら $\mu(N,m)$ は高々 m の 2 倍程度にしかならない。

しかし、もちろん $f(1-0) = +\infty$ なので、m が N の近くならば大きくなってしまうが、元々の問題では、$m > N/2$ の場合は、逆に最初に全部を 1 にしてしまって、$(N-m)$ 個をランダムに 0 にすればよいので、実質的に $m \le N/2$ の場合だけを考えればよく、この場合の期待値 $\mu(N,m)$ と m の比は最大で $f(0.5) = 2\log 2 = 1.39$ 位であることがわかる。

6 $S(k,r)$

この節以降で、3 節で述べた方法での計算を行い、それが 4 節での結果 (3) と一致することを確認する。ただし、ここから先の話は元の問題からすれば本質的な話ではなく、純粋に数学的な議論のみである。また、この方針での計算はかなり大変であるので、この後いくつかの節に分けて考察する。

3 節で考えたように、k 回余計にかかって $(m+k)$ 回で終わる確率は

$$
\begin{aligned}
p_{m+k} &= \frac{N-1}{N}\,\frac{N-2}{N}\cdots\frac{N-m+1}{N}\sum_{1\le i_i\le\ldots\le i_k\le m-1}\frac{i_1}{N}\cdots\frac{i_k}{N} \\
&= \frac{1}{N^m}\begin{pmatrix} N \\ m \end{pmatrix} m!\frac{1}{N^k}S(k,m-1) \quad (1 < m \le N)
\end{aligned}
$$

であるので、期待値 $\mu(N,m)$ は、

$$
\mu(N,m) = \sum_{k=0}^{\infty}(m+k)p_{m+k} = \frac{m!}{N^m}\begin{pmatrix} N \\ m \end{pmatrix}\sum_{k=0}^{\infty}\frac{m+k}{N^k}S(k,m-1) \tag{4}
$$

となる。まずは、この $S(k,r)$ を k と r の式で表すところから始める。なお、$m=1$ のときは明らかに $\mu(N,m) = 1$ であるから、以後 $1 < m \le N$ として考える。

3 節で見たように、$S(k,r)$ には、

$$
S(k,r) = \sum_{i=1}^{r} iS(k-1,i) \tag{5}
$$

$$
S(k,1) = 1 \tag{6}
$$

が成り立つ。$S(1, r) = r(r+1)/2$ も言えていたが、これはむしろ (5) で

$$S(0, r) = 1 \tag{7}$$

であるとすれば得られるので、(7) が成り立つとすればよい。なお、(5) をこの式の r を $(r-1)$ とした式と比較してみればわかるが、

$$S(k, r) = S(k, r-1) + rS(k-1, r) \quad (k \geq 1,\ r \geq 2) \tag{8}$$

が成り立つことに注意する。以後、この (6), (7), (8) から $S(k, r)$ を求めていくことにするが、逆にこれを満たす $S(k, r)$ が一意に決定されることも容易にわかる。

さて、今 (8) で $r = 2$ としてみると、(6) より

$$S(k, 2) = S(k, 1) + 2S(k-1, 2) = 2S(k-1, 2) + 1$$

となるから、この両辺を 2^k で割れば、

$$\frac{S(k, 2)}{2^k} = \frac{S(k-1, 2)}{2^{k-1}} + \frac{1}{2^k}$$

となるので、この式にこの式の k を $(k-1)$ にしたものを代入する、といったことを繰り返すことによって、

$$\begin{aligned}
\frac{S(k, 2)}{2^k} &= \frac{S(k-1, 2)}{2^{k-1}} + \frac{1}{2^k} = \frac{S(k-2, 2)}{2^{k-2}} + \frac{1}{2^{k-1}} + \frac{1}{2^k} = \cdots \\
&= \frac{S(0, 2)}{2^0} + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^k} = 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^k} \\
&= \frac{1 - 1/2^{k+1}}{1 - 1/2} = 2 - \frac{1}{2^k}
\end{aligned}$$

となり、よって

$$S(k, 2) = 2 \cdot 2^k - 1 \tag{9}$$

が得られる。同様に、(8) で $r = 3$ とすると

$$S(k, 3) = S(k, 2) + 3S(k-1, 3) = 3S(k-1, 3) + 2 \cdot 2^k - 1$$

となるので、3^k で割れば

$$\begin{aligned}\frac{S(k,3)}{3^k} &= \frac{S(k-1,3)}{3^{k-1}}+2\left(\frac{2}{3}\right)^k-\left(\frac{1}{3}\right)^k=\frac{S(0,3)}{3^0}+2\sum_{j=1}^{k}\left(\frac{2}{3}\right)^j-\sum_{j=1}^{k}\left(\frac{1}{3}\right)^j\\ &= 1+\frac{4}{3}\,\frac{1-(2/3)^k}{1-2/3}-\frac{1}{3}\,\frac{1-1/3^k}{1-1/3}=1+4-4\left(\frac{2}{3}\right)^k-\frac{1}{2}+\frac{1}{2}\,\frac{1}{3^k}\end{aligned}$$

より、

$$S(k,3)=\frac{9}{2}\cdot 3^k-4\cdot 2^k+\frac{1}{2} \tag{10}$$

のようになる。この、(9), (10) より、

$$S(k,r)=\sum_{j=1}^{r}a_{r,j}j^k=a_{r,1}\cdot 1^k+a_{r,2}\cdot 2^k+\cdots+a_{r,r}\cdot r^k \tag{11}$$

の形であることが想像される。

もし、この (11) を (8) に代入して、(6), (7), (8) を満たすように矛盾なく $a_{r,j}$ を決定できれば、前に述べたようにそれを満たす $S(k,r)$ は一意であるから、それで $S(k,r)$ が得られることになる。

(11) を (8) に代入すると、

$$\sum_{j=1}^{r}a_{r,j}j^k=\sum_{j=1}^{r-1}a_{r-1,j}j^k+r\sum_{j=1}^{r}a_{r,j}j^{k-1}=\sum_{j=1}^{r-1}\left(a_{r-1,j}+\frac{r}{j}a_{r,j}\right)j^k+a_{r,r}r^k$$

となる。よって、$1\le j\le r-1$ に対して、

$$a_{r,j}=a_{r-1,j}+\frac{r}{j}\,a_{r,j}$$

であればよく、これは、

$$a_{r,j}=\frac{-j}{r-j}\,a_{r-1,j}$$

を意味するので、

$$
\begin{aligned}
a_{r,j} &= \frac{-j}{r-j}\,a_{r-1,j} = \frac{-j}{r-j}\,\frac{-j}{r-1-j}\,a_{r-2,j} = \cdots \\
&= \frac{-j}{r-j}\,\frac{-j}{r-1-j}\cdots\frac{-j}{1}\,a_{j,j}
\end{aligned}
$$

となり、

$$
a_{r,j} = \frac{(-j)^{r-j}}{(r-j)!}\,a_{j,j} \tag{12}
$$

となることになる。この $a_{j,j}$ は次のように決定できる。(11) に $k=0$ を代入すると (7) より

$$
S(0,r) = \sum_{j=1}^{r} a_{r,j} = 1
$$

となるが、ここに (12) を代入して

$$
\sum_{j=1}^{r} \frac{(-j)^{r-j}}{(r-j)!} a_{j,j} = 1 \quad (r \geq j) \tag{13}
$$

が得られるが、この (13) から次々 $a_{j,j}$ を求めることができる。例えば、(13) に $r=1$ を代入すれば

$$
\frac{(-1)^0}{0!} a_{1,1} = 1
$$

より $a_{1,1}=1$ となり、$r=2$ を代入すると、

$$
\frac{(-1)^1}{1!} a_{1,1} + \frac{(-2)^0}{0!} a_{2,2} = 1
$$

となるので、$a_{2,2} = a_{1,1} + 1 = 2$ と求まる。以下、計算すると、

$$
a_{3,3} = \frac{9}{2} = \frac{3^2}{2},\ a_{4,4} = \frac{32}{3} = \frac{4^3}{6},\ a_{5,5} = \frac{625}{24} = \frac{5^4}{24}
$$

のようになるので、

$$a_{j,j} = \frac{j^{j-1}}{(j-1)!} = \frac{j^j}{j!} \tag{14}$$

となることが予想される。これは実際に、(13) と帰納法を用いれば証明できる。$j < r$ まで (14) が成り立つとすれば、(13) により $a_{r,r}$ は、

$$a_{r,r} = 1 - \sum_{j=1}^{r-1} \frac{(-j)^{r-j}}{(r-j)!} \frac{j^j}{j!} = 1 - \sum_{j=1}^{r-1} (-1)^{r-j} \begin{pmatrix} r \\ j \end{pmatrix} \frac{j^r}{r!}$$

となる。ここで、

$$\sum_{j=0}^{r} \begin{pmatrix} r \\ j \end{pmatrix} (-1)^{r-j} j^r = r! \tag{15}$$

であることを用いれば、

$$\begin{aligned} a_{r,r} &= 1 - \frac{1}{r!} \left\{ \sum_{j=0}^{r} (-1)^{r-j} \begin{pmatrix} r \\ j \end{pmatrix} j^r - (-1)^0 \begin{pmatrix} r \\ r \end{pmatrix} r^r - 0 \right\} \\ &= 1 - \frac{1}{r!}(r! - r^r) = \frac{r^r}{r!} \end{aligned}$$

となるので、帰納法により (14) が成り立つことになる。

(15) は、母関数の方法を用いて以下のように証明できる。二項定理により、

$$f_1(x) = \sum_{j=0}^{r} \begin{pmatrix} r \\ j \end{pmatrix} (-1)^{r-j} e^{jx} = (e^x - 1)^r$$

であるが、$(e^{jx})^{(r)} = j^r e^{jx}$ なので (15) は $f_1^{(r)}(0)$ に等しいことになる。

一方、テイラー展開を考えると、

$$\begin{aligned} f_1(x) &= (e^x - 1)^r = \left(x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \right)^r = x^r + O(x^{r+1}) \\ & \quad (O(x^{r+1}) \text{ は } (r+1) \text{ 次以上の項の和を意味する}) \end{aligned}$$

となるので $f_1^{(r)}(0)=r!$ となり、これで (15) が示されたことになる。

結局、(11), (12), (14) により、

$$S(k,r)=\sum_{j=1}^{r}\frac{(-j)^{r-j}}{(r-j)!}\frac{j^j}{j!}j^k=\sum_{j=1}^{r}(-1)^{r-j}\begin{pmatrix} r \\ j \end{pmatrix}\frac{j^{r+k}}{r!} \tag{16}$$

となる。これは、$r=1$ とすると

$$S(k,1)=(-1)^0\begin{pmatrix} 1 \\ 1 \end{pmatrix}\frac{1^{1+k}}{1!}=1$$

となり (6) も満たすので、(6), (7), (8) を矛盾なく満たすことになり、確かに (16) が成り立つことがわかる。

7 $\mu(N,m)$

6 節で求めた $S(k,r)$ から $\mu(N,m)$ を計算してみる。(16) を (4) に代入すると、

$$\begin{aligned}\mu(N,m) &= \frac{m!}{N^m}\begin{pmatrix} N \\ m \end{pmatrix}\sum_{k=0}^{\infty}\frac{m+k}{N^k}\sum_{j=1}^{m-1}(-1)^{m-1-j}\begin{pmatrix} m-1 \\ j \end{pmatrix}\frac{j^{m-1+k}}{(m-1)!} \\ &= \frac{m}{N^m}\begin{pmatrix} N \\ m \end{pmatrix}\sum_{j=1}^{m-1}(-1)^{m-1-j}\begin{pmatrix} m-1 \\ j \end{pmatrix}j^{m-1}\sum_{k=0}^{\infty}(m+k)\left(\frac{j}{N}\right)^k\end{aligned}$$

となる。$j/N\le(m-1)/N<1$ であるから、最後の和は収束し、(1), (2) より

$$\begin{aligned}\sum_{k=0}^{\infty}(m+k)\left(\frac{j}{N}\right)^k &= m\sum_{k=0}^{\infty}\left(\frac{j}{N}\right)^k+\sum_{k=0}^{\infty}\left(\frac{j}{N}\right)^{k-1}\frac{j}{N} \\ &= m\frac{1}{1-j/N}+\frac{j}{N}\frac{1}{(1-j/N)^2}=\frac{Nm}{N-j}+\frac{Nj}{(N-j)^2}\end{aligned}$$

となるので、$\mu(N,m)=\mu_\alpha+\mu_\beta$ と分け、

$$\begin{aligned}\mu_\alpha &= \frac{Nm^2}{N^m}\begin{pmatrix} N \\ m \end{pmatrix}\sum_{j=1}^{m-1}(-1)^{m-1-j}\begin{pmatrix} m-1 \\ j \end{pmatrix}\frac{j^{m-1}}{N-j}, \\ \mu_\beta &= \frac{Nm}{N^m}\begin{pmatrix} N \\ m \end{pmatrix}\sum_{j=1}^{m-1}(-1)^{m-1-j}\begin{pmatrix} m-1 \\ j \end{pmatrix}\frac{j^m}{(N-j)^2}\end{aligned}$$

と書けることになる。

まずは μ_α の和の部分を考える。

$$F_1(N,n)=\sum_{j=1}^{n}(-1)^{n-j}\left(\begin{array}{c}n\\j\end{array}\right)\frac{j^n}{N-j}=\sum_{j=0}^{n}\left(\begin{array}{c}n\\j\end{array}\right)(-1)^{n-j}\frac{j^n}{N-j}\quad(1\leq n<N)$$

とし $f_2(x,y)$ を、

$$f_2(x,y)=\sum_{j=0}^{n}\left(\begin{array}{c}n\\j\end{array}\right)(-1)^{n-j}e^{jx}e^{(N-j)y}$$

とすると、二項定理より

$$\begin{aligned}f_2(x,y)&=\sum_{j=0}^{n}\left(\begin{array}{c}n\\j\end{array}\right)(-1)^{n-j}(e^{x-y})^je^{Ny}=(e^{x-y}-1)^ne^{Ny}\\&=(e^x-e^y)^ne^{(N-n)y}\end{aligned}$$

となる。一方、$N-j\geq N-n>0$ なので、

$$\int_{-\infty}^{0}e^{(N-j)y}dy=\left[\frac{1}{N-j}e^{(N-j)y}\right]_{y=-\infty}^{y=0}=\frac{1}{N-j}$$

が成り立つから、

$$f_3(x)=\int_{-\infty}^{0}f_2(x,y)dy$$

とすれば、$F_1(N,n)=f_3^{(n)}(0)$ となることになる。

$f_3(x)$ の積分を $e^y=t$ と置換して計算すると、

$$f_3(x)=\int_{-\infty}^{0}(e^x-e^y)^ne^{(N-n)y}dy=\int_0^1(e^x-t)^nt^{N-n-1}dt$$

となるが、これをさらに $t=e^xs$ と置換すると、

$$f_3(x)=e^{Nx}\int_0^{e^{-x}}(1-s)^ns^{N-n-1}ds$$

となる。この積分の部分を $f_4(x)$ とすると、$f_4'(x)$ のテイラー展開は

$$
\begin{aligned}
f_4'(x) &= \left(\int_0^{e^{-x}} (1-s)^n s^{N-n-1} ds\right)' = -e^{-x}(1-e^{-x})^n e^{(-N+n+1)x} \\
&= -e^{-Nx}(e^x-1)^n = -(1+O(1))(x+O(x^2))^n \\
&= -x^n + O(x^{n+1})
\end{aligned}
$$

となるので、これを積分すれば

$$
f_4(x) = f_4(0) - \frac{x^{n+1}}{n+1} + O(x^{n+2}) \tag{17}
$$

であることがわかる。よって $f_3(x)$ は

$$
f_3(x) = f_4(0)e^{Nx} + O(x^{n+1})
$$

となり、よって $F_1(N,n)$ は

$$
F_1(N,n) = f_3^{(n)}(0) = N^n f_4(0)
$$

と表される。ここで $f_4(0)$ はベータ関数で書け、

$$
\begin{aligned}
f_4(0) &= \int_0^1 (1-s)^n s^{N-n-1} ds = B(n+1, N-n) = \frac{\mathit{\Gamma}(n+1)\mathit{\Gamma}(N-n)}{\mathit{\Gamma}(N+1)} \\
&= \frac{n!(N-n-1)!}{N!} = \frac{(n+1)!(N-n-1)!}{(n+1)N!} = \frac{1}{(n+1)\begin{pmatrix} N \\ n+1 \end{pmatrix}}
\end{aligned}
$$

となり、結局 $F_1(N,n)$ は

$$
F_1(N,n) = \frac{N^n}{(n+1)\begin{pmatrix} N \\ n+1 \end{pmatrix}}
$$

と書けることになる。これにより μ_α は

$$
\mu_\alpha = \frac{m^2}{N^{m-1}}\begin{pmatrix} N \\ m \end{pmatrix} F_1(N,m-1) = \frac{m^2}{N^{m-1}}\begin{pmatrix} N \\ m \end{pmatrix} \frac{N^{m-1}}{m\begin{pmatrix} N \\ m \end{pmatrix}} = m \tag{18}
$$

となる。次に μ_β の和の部分

$$F_2(N,n)=\sum_{j=1}^{n}(-1)^{n-j}\left(\begin{array}{c}n\\ j\end{array}\right)\frac{j^{n+1}}{(N-j)^2}=\sum_{j=0}^{n}\left(\begin{array}{c}n\\ j\end{array}\right)(-1)^{n-j}\frac{j^{n+1}}{(N-j)^2}$$

を考える $(1\le n<N)$。この場合、

$$\begin{aligned}\int_{-\infty}^{0}\int_{-\infty}^{t}e^{(N-j)y}dydt &= \int_{-\infty}^{0}\left[\frac{1}{N-j}e^{(N-j)y}\right]_{y=-\infty}^{y=t}dt=\frac{1}{N-j}\int_{-\infty}^{0}e^{(N-j)t}dt\\ &= \frac{1}{(N-j)^2}\end{aligned}$$

を利用して、

$$f_5(x)=\int_{-\infty}^{0}\int_{-\infty}^{t}f_2(x,y)dydt=\sum_{j=1}^{n}\left(\begin{array}{c}n\\ j\end{array}\right)(-1)^{n-j}\frac{e^{jx}}{(N-j)^2}$$

とすれば、$F_2(N,n)=f_5^{(n+1)}(0)$ となることがわかる。一方 $f_3(x)$ と同様の計算により $f_5(x)$ は

$$\begin{aligned}f_5(x) &= \int_{-\infty}^{0}\int_{y}^{0}f_2(x,y)dtdy=-\int_{-\infty}^{0}yf_2(x,y)dy\\ &= -\int_{-\infty}^{0}(e^x-e^y)^n e^{(N-n)y}ydy=-\int_{0}^{1}(e^x-t)^n t^{N-n-1}\log tdt\\ &= -e^{Nx}\int_{0}^{e^{-x}}(1-s)^n s^{N-n-1}(\log s+x)ds\\ &= -e^{Nx}f_6(x)-xf_3(x)\end{aligned}$$

と変形できる。ここで $f_6(x)$ は

$$f_6(x)=\int_{0}^{e^{-x}}(1-s)^n s^{N-n-1}\log sds$$

とする。$f_6'(x)$ のテイラー展開は、

$$f_6'(x)=-e^{-x}(1-e^{-x})^n e^{(-N+n+1)x}(-x)=e^{-Nx}(e^x-1)^n x=x^{n+1}+O(x^{n+2})$$

となるので、これを積分すれば

$$f_6(x) = f_6(0) + \frac{x^{n+2}}{n+2} + O(x^{n+3}) \tag{19}$$

となるから、$f_5(x)$ は (17), (19) より

$$\begin{aligned} f_5(x) &= -e^{Nx}\{f_6(0) + O(x^{n+2})\} - f_4(0)xe^{Nx} + O(x^{n+2}) \\ &= -f_6(0)e^{Nx} - f_4(0)xe^{Nx} + O(x^{n+2}) \end{aligned}$$

となるので、よって $F_2(N, n)$ は

$$\begin{aligned} F_2(N, n) &= f_5^{(n+1)}(0) = -N^{n+1}f_6(0) - (n+1)N^n f_4(0) \\ &= -N^{n+1}f_6(0) - \frac{N^n}{\begin{pmatrix} N \\ n+1 \end{pmatrix}} \end{aligned}$$

と表されることになる。よって μ_β は、

$$\begin{aligned} \mu_\beta &= \frac{m}{N^{m-1}}\begin{pmatrix} N \\ m \end{pmatrix} F_2(N, m-1) = \frac{m}{N^{m-1}}\begin{pmatrix} N \\ m \end{pmatrix}\left\{-N^m f_6(0) - \frac{N^{m-1}}{\begin{pmatrix} N \\ m \end{pmatrix}}\right\} \\ &= -Nm\begin{pmatrix} N \\ m \end{pmatrix} f_6(0) - m \end{aligned}$$

となり、(18) より、

$$\mu(N, m) = \mu_\alpha + \mu_\beta = -Nm\begin{pmatrix} N \\ m \end{pmatrix} f_6(0) \tag{20}$$

となるので、結局この

$$f_6(0) = \int_0^1 (1-s)^n s^{N-n-1} \log s ds \quad (n = m-1)$$

を求めればいいことになる。

8 $f_6(0)$

この節では、7 節で考察した $f_6(0)$ を求める。以後この $f_6(0)$ を

$$F_3(N,n)=\int_0^1(1-s)^n s^{N-n-1}\log s ds \tag{21}$$

のようにおいてこれを考えることにする。なお、$m\geq 2\leq N$ で $n=m-1$ なので、ここまでは $1\leq n<N$ としてきたが、この (21) を考える場合は $n=0$ を除外する必要はないので、この (21) は $0\leq n<N$ として考えることにする。

部分積分により、

$$\begin{aligned}
F_3(N,n) &= \int_0^1(1-s)^n\left(\frac{s^{N-n}}{N-n}\right)'\log s ds\\
&= \left[(1-s)^n\frac{s^{N-n}}{N-n}\log s\right]_{s=0}^{s=1}-\int_0^1\frac{s^{N-n}}{N-n}\left\{(1-s)^n\log s\right\}' ds\\
&= -\frac{1}{N-n}\int_0^1 s^{N-n}\left\{(-n)(1-s)^{n-1}\log s+(1-s)^n\frac{1}{s}\right\}ds\\
&= \frac{n}{N-n}\int_0^1(1-s)^{n-1}s^{N-n}\log s ds-\frac{1}{N-n}\int_0^1(1-s)^n s^{N-n-1}ds\\
&= \frac{n}{N-n}F_3(N,n-1)-\frac{1}{N-n}B(n+1,N-n)
\end{aligned}$$

となるが、この第 2 項は、既に 7 節の $f_4(0)$ の計算で見たように、

$$\frac{1}{N-n}B(n+1,N-n)=\frac{1}{(N-n)(n+1)\begin{pmatrix}N\\n+1\end{pmatrix}}$$

と書けるので、

$$F_3(N,n)=\frac{n}{N-n}F_3(N,n-1)-\frac{1}{(N-n)(n+1)\begin{pmatrix}N\\n+1\end{pmatrix}}$$

となり、両辺を $(N-n)/n$ 倍すれば、

$$\frac{N-n}{n}F_3(N,n)=F_3(N,n-1)-\frac{1}{n(n+1)\begin{pmatrix}N\\n+1\end{pmatrix}} \tag{22}$$

となる。ここで、

$$\begin{aligned}\binom{N-1}{n-1}\frac{N-n}{n} &= \frac{(N-1)(N-2)\cdots(N-n+1)}{(n-1)!}\,\frac{N-n}{n}\\ &= \binom{N-1}{n}\end{aligned}$$

なので、(22) の両辺を $\binom{N-1}{n-1}$ 倍すれば

$$\binom{N-1}{n}F_3(N,n)=\binom{N-1}{n-1}F_3(N,n-1)-\frac{\binom{N-1}{n-1}}{n(n+1)\binom{N}{n+1}}$$

と書ける。この最後の項は、

$$\frac{\binom{N-1}{n-1}}{n(n+1)\binom{N}{n+1}}=\frac{(N-1)(N-2)\cdots(N-n+1)}{N(N-1)\cdots(N-n)}=\frac{1}{N(N-n)}$$

となるので、結局

$$\binom{N-1}{n}F_3(N,n)=\binom{N-1}{n-1}F_3(N,n-1)-\frac{1}{N(N-n)}$$

が成り立つことになる。この右辺の最初の項は、左辺の n を $(n-1)$ にした形になっているので繰り返し代入すれば

$$\begin{aligned}&\binom{N-1}{n}F_3(N,n)\\ &= \binom{N-1}{n-2}F_3(N,n-2)-\frac{1}{N(N-n+1)}-\frac{1}{N(N-n)}=\cdots\\ &= \binom{N-1}{0}F_3(N,0)-\left\{\frac{1}{N(N-1)}+\frac{1}{N(N-2)}+\cdots+\frac{1}{N(N-n)}\right\}\end{aligned}$$

が得られる。ここで、

$$F_3(N,0)=\int_0^1 s^{N-1}\log s ds=\left[\frac{s^N}{N}\log s\right]_{s=0}^{s=1}-\int_0^1\frac{s^N}{N}\frac{1}{s}ds=-\left[\frac{s^N}{N^2}\right]_{s=0}^{s=1}=-\frac{1}{N^2}$$

なので、よって

$$\binom{N-1}{n}F_3(N,n)=-\frac{1}{N}\sum_{j=0}^{n}\frac{1}{N-j}$$

となり、結局 $F_3(N,n)$ は

$$F_3(N,n)=\frac{-1}{N\binom{N-1}{n}}\sum_{j=0}^{n}\frac{1}{N-j}=\frac{-1}{(n+1)\binom{N}{n+1}}\sum_{j=0}^{n}\frac{1}{N-j}$$

と表されることになる。よって (20) より、

$$\begin{aligned}\mu(N,m) &= -Nm\binom{N}{m}F_3(N,m-1)\\ &= -Nm\binom{N}{m}\left\{\frac{-1}{m\binom{N}{m}}\sum_{j=0}^{m-1}\frac{1}{N-j}\right\}=\sum_{j=0}^{m-1}\frac{N}{N-j}\end{aligned}$$

となり、これでようやく (3) と同じものが得られたことになる。

参考文献

[1] 竹野茂治「プログラム中のある乱数の評価」(2007 年 5 月)