# 誕生日一致確率の多項式評価

An estimate on coincidence probabilities of birthdays by polynomials

中 村 宗 敬
Munetaka NAKAMURA

山梨大学教育人間科学部紀要　第 17 巻　2015年度抜刷

平成 27 年（2015年）度　　山梨大学教育人間科学部紀要　　第 17 巻　P 175～181

# 誕生日一致確率の多項式評価

An estimate on coincidence probabilities of birthdays by polynomials

中 村 宗 敬

Munetaka NAKAMURA

概要

誕生日問題における一致確率を，対象人数$n$に関する二項係数を用いて$N^{-1}=1/365$の多項式として表現し，そこから得られる近似式が$n$が小さいときは精度が良いことを示す。また多項表現の係数の具体的な表現を探る。

## 1. はじめに

今日「誕生日問題」と呼ばれているものは，おそらく歴史的に早い時期から考えられたのであろうが，形に残る明確な起源は，頻度的確率論におけるコレクティフ概念の導入者として有名なフォン・ミーゼス（1883－1953）のようである（[1]，[4]）。60人の集団の中に3人の同じ誕生日の人が見つかって，これを滅多に起こらない珍しいことだと驚いた人が，この確率はどれくらいかとフォン・ミーゼスに尋ねたのが発端とのことである。

時は移って，現在は「ある人数の集団中に同じ誕生日の人が現れる確率はどれくらいか」とか，特に，「その確率が 1/2 を超えるには少なくとも何人が必要か」という問題に主眼点が移っている。一般向けに確率論を解説した書籍や高等学校の教科書等で，この問題に言及される時もこの形である。当初の問題のように3人同じというのはなかなか起こらないだろうが（後述するように，このためには100前後の人数が必要である），多くの人が考えるよりも少ない人数で一致ペアは現れる。しかも，それが実際に簡単に確認できるということで興味が惹かれるのであろう。例えば，40人のクラスで同じ誕生日の人がいる確率は 0.9 を超える。高校などのクラスで確認すると，文字通り十中八九誕生日一致ペアがいることになる。

以下では，上に言及した現在の形のうち前者，すなわち，$n$人の集団の中で同じ誕生日の人が出現する確率はどれくらいか，という問題にまずは焦点をあてて考察をしてみる。後者に関してはそこから自ずと答えを導き出せる。より広い占有問題の一般的取り扱いに対しては，[1] 等を参照されたい。

まず，誕生日問題のモデルとして，1 年の日数を365とし誕生日の分布は一様分布を仮定して考察ことにする。そこで，まず $n$ 人の中に同じ誕生日の人がいない確率を $q_n$ としてみよう。すると，簡単にわかるように，これは $n$ 人の誕生日がすべ異なると言い換えることができるから，

$$q_n=\frac{(N)_n}{N_n}=\prod_{i=0}^{n-1}\left(1-\frac{i}{N}\right)$$

となる。ただし，$N=365$，$(N)_n=\prod_{i=0}^{n-1}(N-i)$ である。$(N)_n$は組合せ論的記号を用いれば，上に述べたとおりの意味の ${}_NP_n$ であるが，後に述べるスターリング数との関連を考慮してこの記法を用いることにする。

したがって，$n$ 人の中に同じ誕生日の人がいる確率 $P_n$は

$$P_n=1-q_n=1=\prod_{i=0}^{n-1}\left(1-\frac{i}{N}\right)$$

になる。これは比較的よく知られた結果であろう。これによって $n$ を増加させて順に数値計算していくと，思いのほか $p_n$ の値が大きいというのが，誕生日問題がパラドックスとして引用される理由である。具体的に数値例をあげると，$p_{10}=0.1233$，$p_{20}=0.4114$，$p_{30}=0.7063$，$p_{40}=0.8912$ 等々（$n=23$ で 1/2 を超える）。

現在の PC 等の計算機の普及状況からすれば，これらの数値計算も何ら困難を来さないであろうし，かなりの精度の良いポアソン近似式 $p_n \approx 1-\exp(-n(n-1)/(2N))$ も利用できるが, より素朴に（固定された$n$に関して）$N^{-1}$の多項式として $p_n$ の近似式を見つけるのが本論の主旨である。

## 2. 第１種スターリング数による表現

上述の誕生日一致確率に関して第１種スターリング数を利用するのが簡明であろう。スターリング数についての詳細は, 例えば [2], [3] を参照されたい。$x$を実数として$(x)_n := \prod_{i=0}^{n-1}(x-i)$ のように記す。$(x)_n$ は $x$ の $n$ 次の多項式となるから，これを展開した式の中での $x^k$ の係数を $(-1)^{n-k}\begin{bmatrix} n \\ k \end{bmatrix}$ と書くことにする；$(x)_n = \sum_{k=1}^{n}(-1)^{n-k}\begin{bmatrix} n \\ k \end{bmatrix}x^k.$

このとき，

$$q_n=\frac{(N)_n}{N^n}=N^{-n}\sum_{k=1}^{n}(-1)^{n-k}\begin{bmatrix} n \\ k \end{bmatrix}n^k=\sum_{k=1}^{n}(-1)^{n-k}\begin{bmatrix} n \\ k \end{bmatrix}N^{-(n-k)}=\sum_{l=0}^{n-1}(-1)^l\begin{bmatrix} n \\ k-l \end{bmatrix}N^{-l}=1+\sum_{l=1}^{n-1}(-1)^l\begin{bmatrix} n \\ n-l \end{bmatrix}N^{-l}$$

である。最後の式の冒頭の 1 は$l=0$の項 $(-1)^0\begin{bmatrix} n \\ n-0 \end{bmatrix}N^{-0}$ を別に抜き出して書いたものである。

$\begin{bmatrix} n \\ n-0 \end{bmatrix}=1$ は定義より明らかであろう。これより

$$p_n=-\sum_{l=1}^{n-1}(-1)^l\begin{bmatrix} n \\ n-l \end{bmatrix}N^{-l}=\begin{bmatrix} n \\ n-1 \end{bmatrix}N^{-1}-\begin{bmatrix} n \\ n-2 \end{bmatrix}N^{-2}+\cdots$$

が得られる。右辺においては$l$が小さい項ほど大きさを支配するので，$p_n$ の大きさを見積もるには，$\begin{bmatrix} n \\ n-1 \end{bmatrix}, \begin{bmatrix} n \\ n-2 \end{bmatrix}, \cdots$の順にわかればよい。 ところで，定義より漸化式

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix} = +\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}\quad (1\leq k\leq n)$$

が容易に見て取れる。$\begin{bmatrix} n \\ 0 \end{bmatrix}=0$ に注意する。証明は例えば[2] を参照のこと。これにより, 第１種スターリング数の三角形を作ることができる。縦の1,2,⋯が $n$ を，横の1,2,⋯が $k$ を，交叉する区画の数が $\begin{bmatrix} n \\ k \end{bmatrix}$ を表す。

表１では，計算は上の行から下の行へ，左の列から右の列へと進めていく。すなわち，直前の上の行の左斜め上の数と，すぐ上の数にその行番号の数をかけたものの和を次々に書き加えていけばよい。例えば，$\begin{bmatrix} 6 \\ 3 \end{bmatrix}=5\times\begin{bmatrix} 5 \\ 3 \end{bmatrix}+\begin{bmatrix} 5 \\ 2 \end{bmatrix}=5\times 35+50=225$のように順次計算する。しかし，これを見ても分かる

通り，得られるスターリング数の順序は我々が望む $p_n$ の中の項の大きさの順とは逆であり，かつスターリング数の一般的な形も得られず，あまり有効ではなさそうである。そこで次節以降では少し工夫を試みる。

表１　第１種スターリン数表（縦が $n$, 横が $k$ を表す）

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | |
| 2 | 1 | 1 | | | | | | |
| 3 | 2 | 3 | 1 | | | | | |
| 4 | 6 | 11 | 6 | 1 | | | | |
| 5 | 24 | 50 | 35 | 10 | 1 | | | |
| 6 | 120 | 274 | 225 | 85 | 15 | 1 | | |
| 7 | 720 | 1764 | 1624 | 735 | 175 | 21 | 1 | |
| 8 | 5040 | 13068 | 13132 | 6769 | 1960 | 322 | 28 | 1 |

## 3.　第１種スターリング数の二項係数による表現と近似

この節では，二項係数により第１種スターリング数を表現することを考えることにする。まず，$(x)_n$ を展開して定義から

$$\begin{bmatrix} n \\ k \end{bmatrix} = \sum_{1\le i_1< i_2<\dots< i_{n-k}\le n-1} i_1 i_2 \dots i_{n-k} \quad (\text{和は } 1\le i_1< i_2<\dots< i_{n-k}\le n-1 \text{ を満たす } i_1, i_2, \dots, i_{n-k} \text{ にわたる})$$

であるから，２節の表現に合わせて書き換えて

$$\begin{bmatrix} n \\ n-k \end{bmatrix} = \sum_{1\le i_1< i_2<\dots< i_k\le n-k-1} i_1 i_2 \dots i_k \quad (\text{和は } 1\le i_1< i_2<\dots< i_k\le n-k-1 \text{ を満たす } i_1, i_2, \dots, i_k \text{ にわたる})$$

となることに注意する。これを二項係数によって表現することを試みるのであるが，これに関する次の定理が本論の主要結果である。

**定理 1.**　$\{a(k,j)\}$ を

$a(1,2)=1$，$j\notin\{k+1, k+2,\dots,2k\}$ のとき，$a(k.j)=0$，$a(k+1,j)=(j-1)(a(k.j)+a(k,j-2))$

により再帰的に定義される二重数列とする。このとき，

$$\begin{bmatrix} n \\ n-k \end{bmatrix} = \sum_{**} i_1 i_2 \dots i_k = \sum_{j=k+1}^{2k} a(k.j) \begin{pmatrix} n \\ j \end{pmatrix}$$

である。ただし，$\begin{pmatrix} n \\ j \end{pmatrix}$ は二項係数を表す；$\begin{pmatrix} n \\ j \end{pmatrix} = \dfrac{n!}{j!(n-j)!}$.

**証明**　左側の等式は上に述べたことである。右側の等式を，固定した各 $n$ に対して，$k$ に関する数学的帰納法により証明する。

まず，$\sum_{**} i_1 = \sum_{l=1}^{n-1} i = n(n-1)/2 = \begin{pmatrix} n \\ 2 \end{pmatrix} = a(1,2) \begin{pmatrix} n \\ 2 \end{pmatrix}$ であるから，$k=1$ のときは確かに成り立つ。

次に $k$ のとき成り立つと仮定すると,

$$
\begin{aligned}
\sum_{1\le i_1<i_2<\dots<i_k<l\le n-1} i_2 \quad i_2\dots i_k l &= \sum_{l=1}^{n-1} l = \sum_{1\le i_1<i_2<\dots<i_k\le l-1} i_1 i_2\dots i_k \\
&= \sum_{l=1}^{n-1} l \sum_{j=k+1}^{2k} a(k,j)\binom{l}{j} \quad (\because \text{帰納法の仮定}) \\
&= \sum_{l=1}^{n-1}\sum_{j=k+1}^{2k} a(k,j)(l+1-1)\binom{l}{j} \quad \left(\because (l+1)\binom{l}{j}(j+1)\binom{l+1}{j+1}\right) \\
&= \sum_{l=1}^{n-1}\sum_{j=k+1}^{2k} a(k,j)\left[(j+1)\binom{l+1}{j+1}-\binom{l}{j}\right] \\
&= \sum_{j=k+1}^{2k} a(k,j)\left[(j+1)\sum_{l=1}^{n-1}\binom{l+1}{j+1}-\sum_{l=1}^{n-1}\binom{l}{j}\right] \\
&= \sum_{j=k+1}^{2k} a(k,j)\left[(j+1)\binom{n+2}{j+2}-\binom{n}{j+1}\right] \\
&= \sum_{j=k+1}^{2k}(j+1)a(k,j)\binom{n+1}{j+2}-\sum_{j=k+1}^{2k} a(k,j)\binom{n}{j+1} \\
&= \sum_{j=k+1}^{2k}(j+1)a(k,j)\binom{n}{j+2}+\sum_{j=k+1}^{2k}(j+1)a(k,j)\binom{n}{j+1}-\sum_{j=k+1}^{2k} a(k,j)\binom{n}{j+1} \\
&= \sum_{j=k+1}^{2k}(j+1)a(k,j)\binom{n}{j+2}+\sum_{j=k+1}^{2k}(j+1)a(k,j)\binom{n}{j+1}-\sum_{j=k+1}^{2k} a(k,j)\binom{n}{j+1} \\
&= \sum_{j=k+1}^{2k}(j+1)a(k,j)\binom{n}{j+2}+\sum_{j=k+1}^{2k} ja(k,j)\binom{n}{j+1} \\
&= \sum_{j=k+3}^{2k+2}(j+1)a(k,j-2)\binom{n}{j}+\sum_{j=k+2}^{2k+1}(j-1)a(k,j-1)\binom{n}{j} \\
&= \sum_{j=k+2}^{2k+2}(j-1)a(k,j-2)+a(k,j-1))\binom{n}{j} \quad (\because a(k.k)=a(k.2k+1)=0) \\
&= \sum_{j=(k+1)+1}^{2(k+1)} a(k+1,j)\binom{n}{j}
\end{aligned}
$$

となり，$k+1$のときも成り立つ。したがって，すべての $k$ について成り立つことがわかる。上の変形中に二項係数に関してよく知られた等式

$$
\sum_{l=1}^{n-1}\binom{l+1}{j+1}=\binom{n+1}{j+2},\quad \sum_{l=1}^{n-1}\binom{l}{j}=\binom{n}{j+1}
$$

を用いていることに留意されたい。　■

定理１にしたがって自明でない $a(k,j)$ を計算していくと,次のようになる。

$k=2$

$a(2,3)=2(a(1,2)+a(1,1))=2(1+0)=2, a(2,4)=3(a(1,3)+a(1,2))=3(0+1)=3.$

$k=3$

$a(3,4)=3(a(2,3)+a(2,2))=3(2+0)=6,\ a(3,5)=4(a(2,4)+a(2,3))=4(3+2)=20,$

$a(3,6)=5(a(2,5)+a(2,4))=5(0+3)=15.$

$k=4$

$a(4,5)=4(a(3,4)+a(3,3))=4(6+0)=24$, $a(4,6)=5(a(3,5)+a(3,4))=5(20+6)=130$,

$a(4,7)=6(a(3,6)+a(3,5))=6(15+20)=210$, $a(4,8)=7(a(3,7)+a(3,6))=7(0+15)=105$.

したがって, $P_n$ を $N^{-4}$ まで近似すると

$$p_n \approx \binom{n}{2}N^{-1}-\left(3\binom{n}{4}+2\binom{n}{3}\right)N^{-2}+\left(15\binom{n}{6}+20\binom{n}{5}+6\binom{n}{4}\right)N^{-3}$$
$$-\left(105\binom{n}{8}+210\binom{n}{7}+130\binom{n}{6}+24\binom{n}{4}\right)N^{-4}$$

と求められる。当然のことながら, 第 1 項 $\binom{n}{2}N^{-1}$は $n$ 人のペア一致確率の大枠を示していることになる。この右辺を $a_n$ として, ポアソン近似 $b_n=1-\exp(-n(n-1)/(2N))$と比較してみると, $n=29$までは$a_n$の精度が高い(多項式では全体にわたる近似にやはり無理がある)。確率 1/2 を超えるとして最小整数としてよく引き合いに出される$n=23$の場合, $p_{23}=0.50730$, $a_{23}=0.50665$, $b_{23}=0.50000$である。

## 4.　係数の表現

この節では, 前節の $a(k,j)$のより具体的な表現を導き出してみる。形を簡便にするために, $j$を$k+j$に置き換えて

$$\sum_{1\leq i_1<i_2<\ldots<l_k\leq n-1} i_1 i_2\ldots i_k=\sum_{j=1}^{k}b(k,j)\binom{n}{k+1}$$

としてみよう。そうすると, $b(k,j)$に対する漸化式は

$$b(k+1,j)=(k+j-1)(b(k,j-1)+b(k,j-2))$$

となり, 境界条件は $b(1,1)=1$, $j\notin\{1,2,\ldots,k\}$のとき $b(k,j)=0$ である。これを順次計算して以下の表を得る。

表 2　$b(k.j)$の数表（縦が $k$, 横が $j$ を表す）

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | |
| 2 | 2 | 3 | | | | | | |
| 3 | 6 | 20 | 15 | | | | | |
| 4 | 24 | 130 | 210 | 105 | | | | |
| 5 | 120 | 924 | 2380 | 2520 | 945 | | | |
| 6 | 720 | 7308 | 26432 | 44100 | 34650 | 10395 | | |
| 7 | 5040 | 64224 | 303660 | 705320 | 866250 | 540540 | 135135 | |
| 8 | 40320 | 623376 | 3678840 | 11098780 | 18858840 | 18288270 | 9459450 | 2027025 |

4 行目までは実質的に前節で既に計算しているが, あらためて表を使った計算方法をみてみよう。計算は各行を左から右へ進めればよい。既に計算した 4 行目は次のように進めることができる。まず 1 列目には仮想的な左上の 0 とすぐ上の 6 を加えて, さらにこれに行番号 4 を乗数としてかけ, $(0+6)\times4=24$を置く。 2 列目には左上の 6 とすぐ上の 20 を加えて, 4 に 1 を加えた 5 を乗数としてかけけ$(6+2)\times5=130$, 3 列目には左上の 20 とすぐ上の 15 を加えて, 5 に 1 を加えた 6 を乗数としてかけ$(20+15)\times6=210$, 4 列目には左上の 15 とすぐ上の仮想的な 0 を加えて, 6 に 1 を加えた 7 を乗数としてかけ, $(15+0)\times7=105$をそれぞれ置く。すなわち, きわめて機械的に, 各交叉区画(アドレスを(4,1)のように

書くことにする）には

　$(4,1)$：$(0+6)\times(4+1-1)=24$,

　$(4,2)$：$(6+20)\times(4+2-1)=130$,

　$(4,3)$：$(20+15)\times(4+3-1)=210$,

　$(4,4)$：$(15+0)\times(4+4-1)=105$

を上の行から取り込んで計算できることになる。アドレス$(k,\ j)$に至った段階で$k+j-1$がかけられていることに着目する。したがって，これを繰り返すと次のような表現式が得られる。まず，表中のアドレス$(k,\ j)$に対するアドレス数を$k+j-1$と定義する。このとき，$(j-1)$個の右斜下降↘と$(k-j)$個の直下降↓からなる，$(1,1)$から$(k,\ j)$へのすべての最短経路上のアドレス数の積の和が$b(k,\ j)$に等しくなる。和の項数は$(j-1)+(k-j)=k-1$個のうちどこを右斜下降とするかで決まるから，$\binom{k-1}{j-1}$個になる。例えば，

$$
\begin{aligned}
b(6,4)=&9\times8\times7\times5\times3\times1+9\times8\times6\times5\times3\times1+9\times8\times6\times4\times3\times1+9\times8\times6\times4\times2\times1\\
&+9\times8\times6\times4\times3\times1+9\times7\times6\times5\times3\times1+9\times7\times6\times4\times3\times1+9\times7\times5\times4\times3\times1\\
&+9\times7\times5\times4\times2\times1+9\times7\times5\times3\times2\times1=44100
\end{aligned}
$$

のようにアドレス数積の$\binom{6-1}{4-1}=\binom{5}{3}=10$(個)の和として表現できる。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | |
| 2 | 2 | 3 | | | | | | |
| 3 | 3 | 4 | 5 | | | | | |
| 4 | 4 | 5 | 6 | 7 | | | | |
| 5 | 5 | 6 | 7 | 8 | 9 | | | |
| 6 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | |
| 8 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

図１A　各アドレス数および積$9\times8\times7\times5\times3\times1$に対応する最短経路

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | |
| 2 | 2 | 3 | | | | | | |
| 3 | 3 | 4 | 5 | | | | | |
| 4 | 4 | 5 | 6 | 7 | | | | |
| 5 | 5 | 6 | 7 | 8 | 9 | | | |
| 6 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | |
| 8 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

図１B　各アドレス数および積$9\times8\times6\times5\times3\times1$に対応する最短経路

上記$b(4,6)$の第１項の積 $9\times8\times7\times5\times3\times1$ は 最短経路

$$(1,1)\to(2,2)\to(3,3)\to(4,4)\to(5,4)\to(6,4)$$

より（図1A），第２項の積 $9\times8\times6\times5\times3\times1$ は最短経路

$$(1,1)\to(2,2)\to(3,3)\to(4,3)\to(5,4)\to(6,4)$$

より得られ(図1B)。他の積も残り８本の最短経路より得られる。

また，これを次のように書くこともできる；

$$b(4,6)=9!\Big(\frac{1}{6\times4\times2}+\frac{1}{7\times4\times2}+\frac{1}{7\times5\times2}+\frac{1}{7\times5\times3}+\frac{1}{7\times5\times2}$$
$$+\frac{1}{8\times4\times2}+\frac{1}{8\times5\times2}+\frac{1}{8\times6\times2}+\frac{1}{8\times6\times3}+\frac{1}{8\times6\times4}\Big).$$

各分母は２から８のうちの３数の積からなり，それらの３数は差が２以上のものからなる。これらの３数は，２から８のうち経路に現れなかったアドレス数と見なすこともできる。

一般化すると次の定理を得る。正確な証明は帰納法によればよい。

**定理 2.**　$a<<b \Leftrightarrow a+1<b$と定義すると，

$$b(k,j)=(k+j-1)!\sum_{2\le i_1<<i_2<<\ldots<<i_{j-1}\le k+j-2}\frac{1}{i_1i_2\ldots i_{j-1}}.$$

定理２から

$$b(k,1)=k!,\ b(k,2)=(k+2)!\sum_{l=2}^{k}\frac{1}{l},\ \cdots,\ b(k,k)=\frac{(2k-1)!}{(2k-2)!!}=(2k-1)!!$$

のように，より具体的な表現が得られる。

## 参考文献


［1］ W. Feller, *An introduction to Probability Theory and its Applications, vol.1 (edition 3)*，Wiley, 1968 .（訳）ハイラー, G. ヴァンナー, (訳) 河田 他,『確率論とその応用１上』(新装版)，丸善出版，2012.

［2］ R. Graham, D. Knuth, O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science,* Addison Wesley, 1994.：(訳) 有沢 他,『コンピュータの数学』，共立出版，2005.

［3］ G. Polya, R.E. Tarjan, D.R. Woods, *Notes on Introductory Combinatorics,* Birkhaeuser , 1983：邦訳 G. ポリア, D.R. ウッズ, R.E. タージャン, (訳) 今宮 淳美,『組合せ論入門』，近代科学社，1986.

［4］ 森口繁一,『応用数学夜話』, 筑摩書房, 2011.