

## エクセルで確率を実感する

世の中には確率でもって起こる現象があり、確率を理解することは大切である。しかし、短時間内に多数回繰り返し起こる現象はあまり見られない。確率的な現象の最もよい例であるサイコロ振りにおいても、どの目の出方もほぼ同じ割合になることを確認するには、極めて多数回の実行を必要とする。他方、確率の計算は、細心の論理的な注意をはらって行わないと、誤りを犯しかねない。そのためもあって、確率を苦手と感じる人が少なくない。このような事情があるので、確率の概念を、実感を伴って理解することはそれほど容易なことではない。しかし、パソコンの乱数発生機能を利用する手がある。ここではエクセルを使って、25人以上（25人および50人）の集団の中に誕生日に重なりが起こる状況を観察し、それと計算によって求めた確率の値を見ることによって、確率への関心を誘いたい。1年は365日であるから25人程度の集団の中に誕生日の重なりは起こらないのではないかという予想に反して、その確率は50パーセントを超えるのである。その確率は誕生日が一樣であることをもとに計算しているが、一樣でなく偏りがあれば確率はもっと大きくなるのが、微積分学の多変数関数の極値問題を解くことによって得られる。添付のエクセルファイルのシートは、「25人」、「50人」、「確率」の3つである。

1, エクセルファイルのシート「25人」で25人の誕生日に重なりが起こるかどうかを観察する。

A列に25人の番号を入れ、B列のセルに「=INT(RAND()\*365+1)」を入力することにより、25人それぞれの誕生日（1から365までの通し番号）をランダムにしかも一樣に割り振る。ここで、

**ランダム**とは規則性が無く、何が出てくるか予想がつかないことであり、**一樣**とはどこにも偏りが無いことである。

「RAND()」により、0より大きく1より小さい数がランダム一樣に出てくる。それに365をかけ1を加えた「RAND()\*365+1」により、1より大きく366より小さい数がランダム一樣に出てくる。「INT()」により小数部分を切り捨てるので、「=INT(RAND()\*365+1)」を入力することにより、1から365までの自然数がランダム一樣に出てくるのである。

第2行には25人の番号が、第3行にはB列と同じ誕生日を貼り付けている。

文字「重」が出ているセルの列の番号nと行の番号mの誕生日は一致している。

ただし、 $n = m$  のとき一致するのは当然だから、対角線のセルはすべて「重」となっている。

誕生日が一致するときだけに「重」が出るのは、セルに「=IF(n=m,"重",")」を入力していることによる。

セル C8 に対角線のセル以外の「重」の組がいくらあるかを数え上げている。

キーボードの一番上の行にある「F9」キーを押すと乱数が入れ替わる。何度でも「F9」を押すことを繰り返してみると、重なりの組が無いこともあれば重なりの組があることもあるが、どちらかと言えば重なりの組があることのほうが多いことが観察できる。

2, シート「50人」で50人の誕生日に重なりが起るかどうかを観察する。

同じように50人の場合を観察すると、ほとんどの場合に対角線セル以外に「重」が出る、すなわち、ほとんどの場合に50人の誕生日には重なりが起ることが観察できる。

3, シート「確率」で重なりが起る確率を計算する。

A列には人数「 $n$ 」を、B列には「 $365 - n + 1$ 」の値を、C列には「 $\frac{365-n+1}{365}$ 」の値を、D列には上のセルの値に順次「 $\frac{365-n+1}{365}$ 」を掛けた値が出るようにしている。この値が誕生日が重ならない確率である。E列は1からD列の値を引いた値、すなわち、誕生日が重なる確率である。なぜなら、

$$\begin{aligned} \text{2人の誕生日が重ならない確率} &= \frac{\text{2人の誕生日が重ならない場合の個数}}{\text{2人の誕生日の起りえる場合の個数}} = \frac{365 \times 364}{365 \times 365} \\ &= 1 \times \frac{364}{365} = 0.997 \end{aligned}$$

$$\begin{aligned} \text{3人の誕生日が重ならない確率} &= \frac{\text{3人の誕生日が重ならない場合の個数}}{\text{3人の誕生日の起りえる場合の個数}} = \frac{365 \times 364 \times 363}{365 \times 365 \times 365} \\ &= 1 \times \frac{364}{365} \times \frac{363}{365} = 0.991 \end{aligned}$$

$$\begin{aligned} \text{4人の誕生日が重ならない確率} &= \frac{\text{4人の誕生日が重ならない場合の個数}}{\text{4人の誕生日の起りえる場合の個数}} = \frac{365 \times 364 \times 363 \times 362}{365 \times 365 \times 365 \times 365} \\ &= 1 \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} = 0.983 \end{aligned}$$

同様にして

$$\begin{aligned} \text{25人の誕生日が重ならない確率} &= \frac{\text{25人の誕生日が重ならない場合の個数}}{\text{25人の誕生日の起りえる場合の個数}} = \frac{365 \times 364 \times 363 \times 362 \times \dots \times 341}{365 \times 365 \times 365 \times \dots \times 365} \\ &= 1 \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{342}{365} \times \frac{341}{365} = 0.431 \end{aligned}$$

となる。したがって、25人の誕生日に重なりが起る確率は  $1 - 0.431 = 0.569$ 、57%、すなわち、100回のうち57回くらいは重なりが起るので、重なりが起

こることが重なりが起こらないことよりも少しだけ多い。

$$50 \text{ 人の誕生日が重ならない確率} = 1 \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \cdots \times \frac{317}{365} \times \frac{316}{365} = 0.0296$$

となる。したがって、50人の誕生日に重なりが起こる確率は  $1 - 0.0296 = 0.9704$ 、97%、すなわち、100回のうち97回くらいは重なりが起こるのである。なお、50人の場合は誕生日の重なりが4組や5組であることがあることも観察できる。

以上は、誕生日に偏りが無いという仮定のもとで、計算したが、偏りがあれば、重なりが起こる確率はもっと大きくなる。

4, 誕生日が一様でなく偏りがあるとすれば重なる確率がさらに大きくなることの証明

$X$  および  $Y$  を、ともに  $n$  個の値  $a_1, a_2, a_3, \dots, a_n$  をそれぞれ確率  $p_1, p_2, p_3, \dots, p_n$  ( $p_1 + p_2 + p_3 + \dots + p_n = 1$ ) でとる、互いに独立な2つの確率変数とすると、

$$\begin{aligned} P(X = Y) &= P(X = Y = a_1) + P(X = Y = a_2) + P(X = Y = a_3) + \cdots + P(X = Y = a_n) \\ &= P(X = a_1)P(Y = a_1) + P(X = a_2)P(Y = a_2) + \cdots + P(X = a_n)P(Y = a_n) \\ &= p_1^2 + p_2^2 + p_3^2 + \cdots + p_n^2 \end{aligned}$$

となる。 $P(X = Y)$  は  $n$  個の変数  $p_1, p_2, p_3, \dots, p_n$  についての連続関数であるから、最小値をもつ。最小値をとる点は極値をとる点であるから、条件  $p_1 + p_2 + p_3 + \cdots + p_n = 1$  のもとでの  $n$  変数関数  $p_1^2 + p_2^2 + p_3^2 + \cdots + p_n^2$  の条件付き極値問題を解けばよい。ここでは  $n-1$  変数関数  $f(p_1, p_2, \dots, p_{n-1}) = p_1^2 + p_2^2 + \cdots + (1 - p_1 - p_2 - \cdots - p_{n-1})^2$  の(条件付きではない)極値問題として解く。極値をとる点において、偏導関数の値は0であるから、

$$\begin{aligned} \frac{\partial f}{\partial p_1} &= 2p_1 - 2(1 - p_1 - p_2 - \cdots - p_{n-1}) = 2(p_1 - p_n) = 0 \\ \frac{\partial f}{\partial p_2} &= 2p_2 - 2(1 - p_1 - p_2 - \cdots - p_{n-1}) = 2(p_2 - p_n) = 0 \\ &\vdots \\ \frac{\partial f}{\partial p_{n-1}} &= 2p_{n-1} - 2(1 - p_1 - p_2 - \cdots - p_{n-1}) = 2(p_{n-1} - p_n) = 0 \end{aligned}$$

とする。これより、 $p_1 = p_2 = p_3 = \cdots = p_{n-1} = p_n = \frac{1}{n}$  を得る。したがって、極値をとるのは等確率(一様)のときである。実際、変数変換

$$p_1 = \frac{1}{n} + x_1, p_2 = \frac{1}{n} + x_2, \dots, p_n = \frac{1}{n} + x_n$$

を行なうと、 $x_1 + x_2 + \cdots + x_n = 0$  だから、

$$\begin{aligned} P(X = Y) &= \left(\frac{1}{n} + x_1\right)^2 + \left(\frac{1}{n} + x_2\right)^2 + \cdots + \left(\frac{1}{n} + x_n\right)^2 \\ &= \left(\frac{1}{n^2} + \frac{2x_1}{n} + x_1^2\right) + \left(\frac{1}{n^2} + \frac{2x_2}{n} + x_2^2\right) + \cdots + \left(\frac{1}{n^2} + \frac{2x_n}{n} + x_n^2\right) \\ &= \frac{1}{n} + x_1^2 + x_2^2 + \cdots + (x_1 + x_2 + \cdots + x_{n-1} + x_n)^2 > \frac{1}{n} \\ &\quad (x_1, x_2, \dots, x_{n-1} \text{のいずれかが } 0 \text{ でないとき}) \end{aligned}$$

だから、等確率（一様）の点からずれると、確率はもっと大きくなる。