

非線形解析による近似の有効性

東京理科大学大学院理工学研究科情報科学専攻 児玉賢史 (Satoshi Kodama)
Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science

1 研究の経緯

演算器は大規模な数値処理を行う際によく利用されているが、非常に大きな値や小さい値が計算過程において発生した場合、誤差の問題がしばしば生じてくる。すなわち、演算機器を用いて何らかの数値計算を行った場合には、常に誤差が発生していないかどうかを判断した上で結果を出力しなければならない。言い換えれば、大規模な数値演算処理において、いくら数学的に正確な値を求めることが可能であったとしても、演算器を通して得られたものは不正確になってしまう場合が非常に多い。

例えばこのような現象の一例として、日常的に利用している一般的な計算機を用いて、任意の正の整数値に関して平方根演算を何度か繰り返した後に、その結果に対して2乗演算を行っていく。本来であれば最初の値に戻るはずだが、出力された演算結果は異なる。これは先に述べたように非常に小さい値や大きい値を『単純に演算器上で計算したため』に発生した誤差によるものである。

一般にそれらの問題を回避する方法として、『有効桁数から実際に必要とされる範囲で(使用目的や利用領域に応じて)求める手法』や『誤差の発生しない別の計算方法を利用する』といった対策をとらなければならない。

しかしながら、前者の場合は(金融工学の分野など)少しのずれも実用上許されない場合等には用いることができず、また、後者の場合は演算過程が複雑になり、結果を出力するまでに多くの過程を要して、最終的に多くの時間やハードといった面でコストが必要となってしまう。しかし、通常利用の範囲(実用上問題のない多くの場合)においては、誤差を許してコストを抑える場合がほとんどである。

そこで今回は単純な計算処理を繰り返し行い、その過程で誤差を発生させていくのではなく、数学的な手法を用いることでおおよその結果を(演算器の得意とする処理法を使用して)求める方法について考察した。

2 非線形近似による確率計算

2.1 誕生日問題に関する演算結果

『60人いれば誕生日が一致している2人が99%以上の確率で存在する』、『50%の確率でよければ23人以上いればよい』といった直感と異なった数学上の一種のパラドックスがある。こ

れは以下のような考察から発生したものである。

一年間を 365 日として、閏年等の特殊な状況を除くとする。この前提のもと n 人が所属するあるグループにおいて『誕生日が重複しない確率 P 』はどの程度かを考える。

1 人であれば、365 日ある誕生日のいずれかの日に必ず生まれているので、確率 $p(1)$ は

$$p(1) = \frac{365}{365}$$

となる。2 人目は 1 人目と異なればよいので確率 $p(2)$ は

$$p(2) = \frac{364}{365}$$

となる。従って一般に、 n 人目は $n - 1$ 人目までと異なるようにすればよいので、確率 $p(n)$ は

$$p(n) = \frac{365 - n + 1}{365}$$

と定義できる。

以上のことから、誕生日が 1 組も一致しない確率 (重複しない確率) $P(n)$ は下記のように計算することができる。

$$P(n) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365 - n + 1}{365}$$

一般に誕生日問題 (誕生日パラドックス) と呼ばれている『誕生日が一致する確率』は上記の $P(n)$ を用いると $1 - P(n)$ で求められる。この問題において n 人の値を変化させて得られた結果を表 1. と図 1. に示す。

表 1. 在籍する人数と重複関係

在籍人数	重複する確率	重複しない確率
1	1.0000000000000000	0.0000000000000000
2	0.9972602739726028	0.002739726027397249
3	0.9917958341152187	0.008204165884781345
4	0.9836440875334498	0.016355912466550215
5	0.9728644263002065	0.027135573699793470
6	0.9595375163508886	0.040462483649111425
7	0.9437642969040246	0.056235703095975365
8	0.9256647076483310	0.074335292351669020
9	0.9053761661108333	0.094623833889166730
10	0.883051822889223	0.116948177711077680
11	0.8588586216782669	0.141141378321733120
12	0.8329752111619355	0.167024788838064500

在籍人数	重複する確率	重複しない確率
13	0.8055897247675705	0.194410275232429490
14	0.7768974879950269	0.223102512004973100
15	0.7470986802363135	0.252901319763686460
16	0.7163959947471499	0.283604005252850100
17	0.6849923347034391	0.315007665296560900
18	0.6530885821282104	0.346911417871789600
19	0.6208814739684630	0.379118526031536950
20	0.5885616164194197	0.411438383580580270
21	0.5563116648347940	0.443688335165206000
22	0.5243046923374497	0.475695307662550300
23	0.4927027656760144	0.507297234323985700
24	0.46165574208547105	0.538344257914529000
25	0.4313002960305360	0.568699703969464000
⋮	⋮	⋮
57	0.009877540658830017	0.990122459341169900
58	0.008335020610738753	0.991664979389261200
59	0.007010551582183006	0.992989448417817000
60	0.005877339134652055	0.994122660865348000

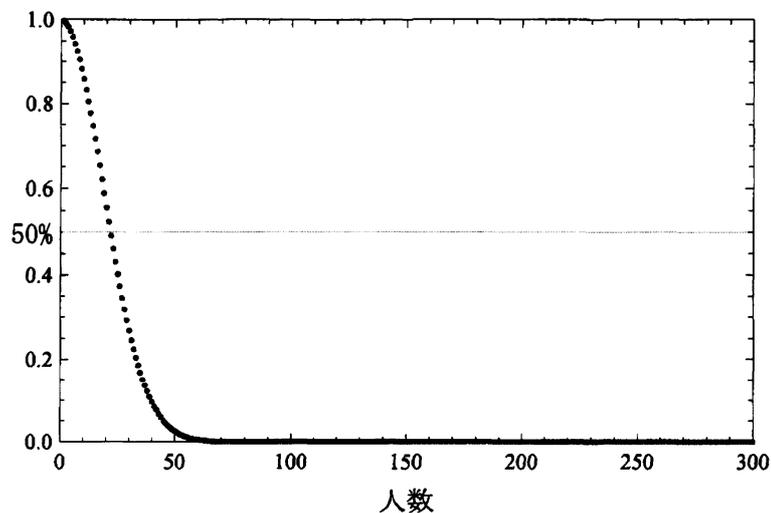


図 1. 在籍する人数と重複しない確率

以上のことから明らかなように、あるグループに属する人が23人以上いれば、そのグループ内に重複する誕生日の人が少なくとも1組以上存在する確率が50%を超えることになる。また、57人いればその確率は99%以上になることがわかる。すなわち、先ほども述べたように、全体的 (n 人中) に同じ誕生日がいる確率は一般的な直感と異なり、かなりの高確率で一致する。

その一方で、補足になるが、特定の人物と同じ誕生日である場合の確率は直感に反することなく、前述した確率よりかなり低くなる。

具体的に論ずると、 n 人が所属するグループで、特定の1人が残り $n-1$ 人のすべてと誕生日が一致しない確率 $P(n)$ は、

$$P(n) = \left(\frac{364}{365}\right)^{n-1}$$

となる。従って、特定の1人と同じ誕生日を持つ人が、残り $n-1$ 人の中に少なくとも1人存在する確率 $\overline{P(n)}$ は

$$\overline{P(n)} = 1 - \left(\frac{364}{365}\right)^{n-1}$$

と計算することが可能である。

表2. にグループに所属する人数と誕生日が異なる確率および重複する確率の関係を示した。

表2. 所属人数と特定人物の誕生日が一致する確率

所属人数	重複する確率	重複しない確率
1	0.0000000000000000	1.0000000000000000
2	0.9972602739726028	0.002739726027397249
3	0.9945280540439106	0.005471945956089352
4	0.9918033196492698	0.008196680350730179
5	0.9890860502803678	0.010913949719632221
6	0.9863762254850791	0.013623774514920917
7	0.9836738248673118	0.016326175132688192
8	0.9809788280868534	0.019021171913146562
9	0.9782912148592182	0.021708785140781783
10	0.9756109649554943	0.024389035044505736
11	0.9729380582021916	0.02706194179780841
12	0.9702724744810897	0.029727525518910336
13	0.9676141937290867	0.03238580627091325
14	0.9649631959380481	0.035036804061951865
15	0.9623194611546563	0.037680538845343725
16	0.95968296948026	0.040317030519740005
17	0.957053701070725	0.04294629892927504
18	0.9544316361362847	0.04556836386371532
19	0.9518167549413907	0.04818324505860927
20	0.949209037804565	0.050790962195435
21	0.9466084650982511	0.05339153490174886
22	0.9440150172486669	0.05598498275133312
23	0.9414286747356568	0.058571325264343166
24	0.9388494180925455	0.06115058190745448

所属人数	重複する確率	重複しない確率
25	0.9362772279059906	0.06372277209400945
26	0.9337120848158372	0.06628791518416277
⋮	⋮	⋮
58	0.8552352523814217	0.14476474761857827
59	0.8528921421009247	0.14710785789907532
60	0.8505554513006481	0.14944454869935186
61	0.8482251623929752	0.15177483760702482
⋮	⋮	⋮
252	0.5022712470827538	0.49772875291724616
253	0.5008951614743079	0.4991048385256921
254	0.4995228459634194	0.5004771540365807
255	0.4981542902210539	0.5018457097789462

表 2. から明らかなように、前回の場合 (条件) と結果が大きく異なり、特定人物の誕生日が一致する確率は 23 人では 6% 程度しかない。なお、この結果は前回の『誕生日が一致する確率が 50% 以上である』という条件を満たすためには少なくとも 253 人より多い人数でなければならないということを示している。これは『特定人物以外に同じ誕生日の組み合わせが何組いても関係ない』という強い条件がはたらいたためである。こちらにも非常に小さい値を使う確率上の問題 (誕生日問題の同列の問題) として追記した。

2.2 誕生日問題に対する近似計算法

先に示した表 1. および図 1. は、 $P(n)$ および $1 - P(n)$ を単純に繰り返して計算した結果を示している。ここでは、先の式を純粋に計算することなく、近似的な結果 (値) を導くことを考える。なお、ここでは比較のために、例としてグループ人数を 60 人として近似的な確率を計算することにした。

1 節で示したように、誕生日が 1 組もない確率 (重複しない確率) P は

$$P = \prod_{k=1}^{60} \left(\frac{365 - k + 1}{365} \right)$$

と表現できる。ここで P に対して \log 演算を適用すると、積演算を和演算に置き換えられ、さらに 365 で割ると以下に示すように刻み幅 $1/365$ の区分求積近似計算が可能になる。

$$\begin{aligned}
\frac{1}{365} \times \log P &= \frac{1}{365} \sum_{k=0}^{59} \log \left(1 - \frac{k}{365} \right) \\
&\approx \int_0^{\frac{60}{365}} \log(1-x) dx \\
&= \left[(x-1) \log(1-x) - x \right]_0^{\frac{60}{365}} \\
&= \left(-\frac{305}{365} \right) \log \left(\frac{305}{365} \right) - \frac{60}{365}
\end{aligned}$$

よって確率 P は次式のようになる。

$$\begin{aligned}
P &= \exp \left(-60 - 305 \log \frac{305}{365} \right) \\
&= 0.005373
\end{aligned}$$

この結果は、先に計算した $n = 60$ (表 1. 参照) の値に非常に近い値となっていることがわかる。
 このような結果が導きだせた理由は、

- ① $1/365$ ずつ減少する階段関数に対して、『非線形関数である対数関数』を用いた近似を行った為
- ② 刻み幅が $1/365$ という小さな値であることに注目して、『階段関数の面積計算』を『階段関数を近似する対数関数の定積分計算』に帰着させた為 (通常は面積を階段関数で近似する)

のようなことからだと考えられる。

一般に、不連続的に変化する関数を連続関数で近似する場合は、線形関数で近似するよりも非線形関数で近似する場合はほうが有効性が高いと考えられている。

3 非線形近似による比率計算

3 節に示した非線形を用いた近似計算には以下のような演算にも適用できると考えられる。

順列と組み合わせに関して、『 n 個のなかから重複を認めないで r 個選び出す場合の数』の比率を求めることを考える。具体的には、

$$\frac{mCr}{nCr} = \frac{m(m-1) \cdots \{m-(r-1)\}}{n(n-1) \cdots \{n-(r-1)\}}$$

を計算することになる。ここで、 $r \leq m \leq n$ を仮定して、 $I(n, m, r)$ を次の式で定義する。

$$I(n, m, r) = \frac{1}{n} \sum_{k=0}^{r-1} \log \left(\frac{m}{n} - \frac{k}{n} \right)$$

前回同様に、 n が十分に大きいとして区分求積法による近似を用いると

$$I(n, m, r) \approx \int_0^{\frac{r}{n}} \log\left(\frac{m}{n} - x\right) dx$$

という式を得ることができる。さらに、 $s \leq t$ を満たす正定数を $s = r/n$ 、 $t = m/n$ と定義すると、 n を十分大きい自然数とした場合、数学的な厳密性を欠くが

$$\begin{aligned} \lim_{n \rightarrow \infty} I(n, nt, ns) &= \int_0^s \log(t - x) dx \\ &= (s - t) \log(t - s) + t \log t - s \end{aligned}$$

が成立する。従って、

$$\frac{1}{n} \log \frac{mCr}{nCr} \approx I(n, m, r) - (n, n, r)$$

という関係が成り立つ。よって、

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\frac{ntCns}{nCns} \right) = (s - t) \log(t - s) + t \log t - (s - 1) \log(1 - s)$$

が成立する。

上記の内容は、例えば、『1 から 90 まで数字が書かれた 90 枚のカードの中から 10 枚を選ぶ場合の数』に対する『1 から 100 までの数字を書いた 100 枚のカードの中から 10 枚を選ぶ場合の数』の比率が求められれば、『1 から 900 まで数字が書かれた 900 枚のカードの中から 100 枚を選ぶ場合の数』に対する『1 から 1000 までの数字を書いた 1000 枚のカードの中から 100 枚を選ぶ場合の数』の比率も近似的に計算可能である事を示している。

ここで、上の式の左辺に関する計算例を表 3. に示す。なお、表 3. の s および t の値は順に 0.1 と 0.9 とした。

表 3. 非線形近似計算を順列と組み合わせに用いた例

n の値	100	200	400	800	1200	2400
$\frac{1}{n} \log \left(\frac{ntCns}{nCns} \right)$	-0.0110722	-0.0111031	-0.0111186	-0.0111263	-0.0111289	-0.0111315

一方、上式の右辺に関する極限值は、

$$(s - t) \log(t - s) + t \log t - (s - 1) \log(1 - s) \approx -0.0111341$$

となるので、数値実験からも、この極限式が成立していることがわかる。

4 考察と発展

先ほど述べたように計算機等を用いて実際に純粋に計算値を求めた場合、通常の数学では考えられない結果が出力される場合がある。例えば前例の『区分求積分を用いた定積分の近似計

算』においては、『刻み幅を小さくするほど近似精度が向上する』と推測できるが、実際にはそういう結果を得られるとは限らない。また1節でも簡単な例を挙げたが、単純な計算においても『非常に大きな値や小さい値を用いた計算』においては、『特定の値以下が演算器上表現できない』といった現象もしばしば発生する。すなわち、計算機などのように物理的に限界のある演算機器を用いて計算を行うには、演算過程においても厳密に評価する必要がある。

このような理由から演算器処理上の誤差を考えた上で結果を得るより、一定の予測可能な数学上の近似を利用する方が実用上有効な場合も多いと考えられる。

演算器は大規模な数値処理において、高速に正確な結果を出力できると思われがちだが、実際には多くの誤差処理を行わなければならない。その際、本文で述べた非線形を用いた近似処理を併用して行うことは有効な手段として発展可能であると考えられる。

本稿で紹介したアイデアは、例えば、ド・モアブル-ラプラスの定理に例を見るように、「2項分布に従う確率変数に、正規化を施して得られる新たなる確率変数の分布は、試行回数が増やしたときに正規分布に法則収束する」という考え方と類似するものである。従って、近似収束先となる非線形関数の形状が把握できていれば、「通常の作業では無限大に発散する可能性のある数値計算を、区分求積法による近似値計算に帰着させるという作業」を適用することによって、非線形近似論的手法を応用することが可能となる。非線形エルゴード定理などは、その代表的なものであると言えるだろう。非線形近似に関する書籍としては文献 [2] を参照することができる。また今回紹介した手法は、「区分求積法で用いる階段関数を対数関数で近似する」という手段を用いているため、当然であるが、「真値との誤差評価」も行わなくてはならない。このような観点から見ると、区間に値を取る関数の積分法も有効になってくると思われる。集合値関数の積分に関する書籍としては、文献 [1] を紹介することができる。

5 参考文献

- [1] Jean-Pierre Aubin and Hérèna Frankowska, Set-Valued Analysis, Birkhäuser, Boston, 1990.
- [2] W. Takahashi, Nonlinear Functional Analysis, Yokohama Publishers, Yokohama, 2000.