付録 C

ちょっとした数学的手法

C.1　比例配分によるデータの内挿

あまり変化が急激でない関数 $y = f(x)$ があって，飛び飛びに関数の値が分かっているものとしよう．そのとき，任意の点の関数値を，それを両側から挟む 2 点のデータから概算することができる．

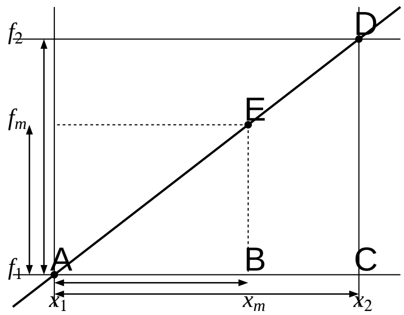

図 C.1　データの比例配分の原理

図 C.1 を見てほしい．斜めの太い直線が，関数 $f(x)$ の一部である．x_1 と x_2 の間は十分に小さいので，$f(x)$ はほとんど直線とみなしてよい．x_1, x_2 における関数の値 $f(x_1)$, $f(x_2)$ をそれぞれ f_1，f_2 とする．x_m はその中間のどこかである．このとき f_m をこの図を使って求めることができる．これを比例配分，あるいは一次の内挿といい，よくある計算手法である．

図をみると三角形 ACD と ABE は相似だから次の式が成り立つ．

$$\frac{x_2 - x_1}{f_2 - f_1} = \frac{x_m - x_1}{f_m - f_1} \tag{C.1}$$

これを変形していけば，次の形が得られる．

$$f_m = f_1 + \frac{(f_2 - f_1)(x_m - x_1)}{x_2 - x_1} \tag{C.2}$$

C.2　有効数字

C.2.1　基本的な考え方

アナログの体重計に乗ってみたところ 52.6 の目盛りと 52.7 の目盛りの真ん中よりやや上のところに針が来た．そこでこれを 52.7 kg として記録した．これは針の位置が 52.7 の目盛りに最も近いから，いいかえれば次の不等式を満たす x を 52.7 としようというのである．

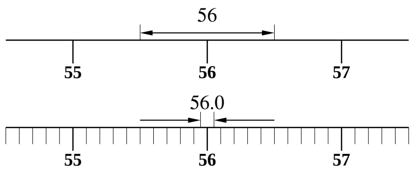

図 C.2　56 と 56.0 のちがい：単に 56 は上の目盛りで読み取った範囲，56.0 なら下の細かい目盛りで読んでいると考えられる．

$$52.65 \leq x < 52.75$$

つまり 52.7 という値が報告された場合には，±0.05 の幅をもつ量であると考えるべきで，とりあえず 0.7 のところは意味のある (significant な) データと見なしてよい．これが四捨五入という操作の意味である．そして意味のある数字は上の桁から 5,2,7 の 3 桁であるから，これを**有効数字または有効桁数を 3 桁である**という．

次に 56.0 kg というデータが記録されていたとしよう．最後の 0 は何のためについているのだろうか．ちょっと考えると 0 があろうとなかろうとゼロはゼロでしかないのだから，これを 56 kg としても何ら不都合はないように思える．しかし，上の有効数字の考えに照らせば，これらはやはり意味が異なるということがわかる．前者は 0.1 kg の桁を四捨五入したものであり，後者は 0.01 kg の桁を四捨五入したものなので，データの精度が大きく異なるのである (図 C.1)．

あるいは実際に即していえば，56.0 というデータを読み取れる秤（はかり）と 56 kg としか読み取れない秤とでは，目盛りの細かさや針の位置の正確さが異なるといってもよい．

C.2.2　実際のデータの有効数字

表 C.1 に，具体的な数値の表記例と，それらの有効桁数の大きさを示した．0.01234 のように，単に大きさを示すためだけに 0 が先頭についている小数については，0 でない数字からが有効数字に含まれる．

表 C.1　有効数字の桁数: *少なくとも 2，多くて 3

表記	有効桁数
12.3	3
12.30	4
0.01234	4
0.0012	2
813	3
810	不明*
810.	3
8.10×10^2	3
6.02×10^{23}	3

810 の場合には，最後の 0 が小数第 1 位を四捨五入して得られたものなのか，あるいは 1 の位が四捨

五入されて 10 の位の 1 が出てきたのかが分からない．前者であれば有効数字は 3 桁であり，後者なら 2 桁ということになる．精度が問題になるようなケースでは，このようなあいまいな表記は好ましくない． ただし 625, 810, 752 などと複数の数値と一緒に並んでいるのなら，推定はできる．

上の問題点を回避するためには指数表記を行うとよい．表にあるように 8.10×10^2 とあれば有効数字は 3 桁，8.1×10^2 とあれば 2 桁と明確に分かる．

C.3　数値の丸め誤差

今，真の値が $a = 0.505$, $b = 1.05$ であるような 2 つの数の積を計算することを考える．

真の積はもちろん，

$$ab = 0.505 \times 1.05 = 0.53025$$

である．ところでこれらをいずれも有効数字 3 桁目で四捨五入したとして，$a' = 0.51$, $b' = 1.1$ と近似値を使って計算すると，

$$a'b' = 0.561$$

となる．この結果は真の値に比べて，

$$\frac{0.561 - 0.53025}{0.53025} = 0.0580$$

すなわち，6% ほども大きすぎる値になっている．このように，四捨五入は誤差を結果に引き込んでしまう処理であるから，計算に使う前の数値を四捨五入するときには慎重な注意を要する．上の場合に即して言うならば，**もし最終結果で有効数字が 2 桁ほしいのであっても，途中計算は 3 桁以上の数値をそのまま使って進めておいて，最後に得られた値だけを四捨五入によって丸めるのが正しい**のである．

C.4　多数回の計算による丸め誤差の蓄積

確率統計の計算では，丸め誤差が発生する要因がかなりある．たとえば次のような計算を見てみよう．二項分布でしばしば現れる次のようなケースはどうだろうか？

$${}_{12}C_4 \times \left(\frac{1}{7}\right)^4 \left(\frac{6}{7}\right)^8 = \frac{12 \cdot 11 \cdot 10 \cdot 9}{4 \cdot 3 \cdot 2 \cdot 1} \left(\frac{1}{7}\right)^4 \left(\frac{6}{7}\right)^8$$

ここで，$1/7 = 0.14285714... \approx 0.143$, $6/7 = 0.85714285... \approx 0.857$ として，ひとつずつ掛け合わせて計算してみると，0.060227... が得られ，一方，可能な限り正確に計算してみると，0.060067... となる．つまり，3 桁に丸めて計算した値は，割合で見ればかなり大きく見積もってしまうことになる．

そこで，$1/7 \approx 0.142857$，$6/7 \approx 0.857412$ で，計算してみると，結果は 0.060066.. と，正確な値にかなり近い．

このように，計算を何度も行う時には，丸めによる誤差がどんどん蓄積することになる[*1]ので，それを避けるためには，なるべく長い桁数で計算を進めていく必要がある．電卓のような簡単な道具で計算を進めるときにも，**多数回の計算による誤差が予想される場合には，目的とする数値の桁数よりも 4 桁以上長い桁数で途中計算を進める**ことが望ましい．

実際の統計用アプリケーションや表計算アプリケーションでは，1 万回を超える繰り返し計算も稀ではないので，ずっと高い精度で内部の計算が行われている．

ただし，**結果の数字のほうは，必要に応じた桁数で表現するべきであって，意味もなく長い桁数で数値を記すのは誤りである．**答を求められている課題の意味を考えて，たとえば日常的な問題で出てくる確率を求めるのであれば，せいぜい 2 桁程度の精度で回答すればよい．

[*1] これも確率論の問題として考えることができ，正規分布が現れるが，ここでは定性的な議論にとどめる．