

コンピュータ・マテリアル・サイエンス 第3回

— コンピュータを使った実験データ解析 (1) —

名古屋工業大学セラミックス基盤工学研究センター 井田 隆

3. 実験誤差の取り扱い

3.1 有効数字

コンピュータは実験データを処理するためにたいへん便利な道具なのですが、コンピュータを使うせいで実験誤差を見損なってしまうことも多いようです。実験データは必ず誤差をともなっているということに注意しなければいけません。

図 3.1 ものさしで棒の長さを測る

たとえば、1 mm 刻みの目盛の物差しで、棒 A の長さを測って 15.8 mm だったとします。このとき、「15.8 mm」という表現自体に、「0.05 mm くらいの誤差はあるかもしれない」という意味が含まれています。ここまでは何も問題はないわけです。

ところが、例えば同じように棒 B の長さを測って、2本の棒の長さの比を求めようとしたときに問題が起きます。棒 B の長さが「10.3 mm」だったとします。

棒 A と B の長さの比を、コンピュータや電卓を使って計算すると例えば「 $15.8 / 10.3 = 1.5339805825$ 」という答が出てきます。そこで、「棒 A と棒 B の長さの比は 1.5339805825 です。」と言う人が少なくないのですが、こう答えると先生から怒られることになります。普通は有効数字 3 桁とみて、「1.53」とするのが正解だとみなされます。

もちろん、今得られている情報からは、「1.53」という数値よりも「1.5339805825」という数値の方が、たとえわずかにせよ「正解に近い確率が高い」と言えなくはありません。それなのにどうして「1.5339805825」が不正解で「1.53」が正解とされるのでしょうか？

それは、一つには、必要な情報を伝えるために、なるべく表現を短くするべきだからということがあります。この場合「1.5339805825」の「1.53」以降の数字はほとんど意味がありません。

もっと重要なことは何も断りなしに「1.5339805825」と言った場合には、誤差が「0.0000000005」くらいであると解釈されるからです。これではまったく実情とは違ってしまっています。誤差を明示しない限り、有効数字が誤差を表現していると自動的に仮定されるということに、注意してください。

有効数字は誤差を考慮に扱うためのもっともコンパクトな表現のしかたです。有効数字を使った計算は、主に以下の2つのルールにしたがいます。

1. 足し算・引き算の結果の有効数字は、最低位の桁が高い位置の方にそろえる。たとえば、

$$1.23 + 0.21325 = 1.44 \quad (1)$$

となります。

2. かけ算・割り算の結果の有効数字の桁数が少ない方にそろえる。たとえば

$$5.32 \times 0.80523 = 4.28 \quad (2)$$

3. 中間結果では1桁以上多く有効数字をとる。

ところで、このような有効数字の計算のルールで本当に納得がいくのでしょうか？上記のルールに従うと、たとえば、

$$4.99/5.000 = 0.998 \quad (3)$$

ですが、

$$5.01/5.000 = 1.00 \quad (4)$$

となってしまうですね。「4.99」という表現は正しい値が4.985から4.995までの範囲にあることを意味していますから、これを5.000で割った値は本当なら0.997から0.999の範囲にあるはずですよ。ところが、有効数字の計算ルールに従った結果「4.99/5.000 = 0.998」と表現すると、これは正しい値が0.9975から0.9985の間にあるという意味になってしまい、本来よりも誤差を過小に評価してしまいます。逆に「5.01/5.000 = 1.00」という表現は、誤差を過大に評価しています。さらに、たとえば実験データの平方根や、指数、対数、三角関数の値を計算した場合の結果がどうなるか、あまりはつきりしません。

3.2 誤差を明示した表現

「4.99」という表現の代わりに、

$$4.990 \pm 0.005 \quad (5)$$

とか、これと同じ意味の

$$4.990(5) \quad (6)$$

という表現を使うことにすれば、単に有効数字を使った場合よりもだいぶ状況が改善されます。

たとえば、「4.990(5)/5.0000(5)」の取りうる最小と最大の値は、

$$(4.990 - 0.005)/(5.0000 + 0.0005) = 0.9969 \quad (7)$$

$$(4.990 + 0.005)/(5.0000 - 0.0005) = 0.9991 \quad (8)$$

から、

$$4.990(5)/5.0000(5) = 0.9980(11) \quad (9)$$

と書けます。

この考え方は、指数や対数などを取る場合にも使えます。たとえば $\ln[4.990(5)]$ の結果の最小と最大は

$$\ln(4.990 - 0.005) = 1.60643 \quad (10)$$

$$\ln(4.990 + 0.005) = 1.60844 \quad (11)$$

ですから、

$$\ln[4.990(5)] = 1.6074(10) \quad (12)$$

と書けるわけです。

このように、結果のとりうる最小と最大を求める方法は、誤差を評価する方法としてはもっとも無難な方法です。しかし、この方法は誤差を過大に評価する傾向があります。

3.3 平均と標準偏差による表現

「4.990(5)」という表現は、とりうる値の範囲を誤差として表しているとも考えられますが、4.990が平均値、誤差 0.005 は標準偏差の意味であるとみなして使うこともできます。

3.3.1 足し算と引き算

変数 x, y が $\langle x \rangle, \langle y \rangle$, 標準偏差が $\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$, $\sigma_y = \sqrt{\langle y^2 \rangle - \langle y \rangle^2}$ (分散 σ_x^2, σ_y^2) のような分布に従う場合, $z = x \pm y$ の平均と分散はどうなるでしょうか?

まず平均は,

$$\begin{aligned}\langle z \rangle &= \langle x \pm y \rangle \\ &= \langle x \rangle \pm \langle y \rangle\end{aligned}\tag{13}$$

となります。

分散は,

$$\begin{aligned}\sigma_z^2 &= \langle (z - \langle z \rangle)^2 \rangle \\ &= \langle z^2 \rangle - \langle z \rangle^2 \\ &= \langle (x \pm y)^2 \rangle - \langle x \pm y \rangle^2 \\ &= \langle x^2 \pm 2xy + y^2 \rangle - (\langle x \rangle \pm \langle y \rangle)^2 \\ &= \langle x^2 \rangle \pm 2\langle x \rangle \langle y \rangle + \langle y^2 \rangle - \langle x \rangle^2 \mp 2\langle x \rangle \langle y \rangle - \langle y \rangle^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2 + \langle y^2 \rangle - \langle y \rangle^2 \\ &= \sigma_x^2 + \sigma_y^2\end{aligned}\tag{14}$$

標準偏差は $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$ となります。つまり、「誤差をともなった数値」どうしの和や差のともなう誤差は、もとの誤差の二乗の和の平方根であるとみなされます。

3.3.2 かけ算

二つの変数 x, y の積 $z = xy$ の平均は,

$$\begin{aligned}\langle z \rangle &= \langle xy \rangle \\ &= \langle x \rangle \langle y \rangle\end{aligned}\tag{15}$$

となり, 分散は

$$\begin{aligned}\sigma_z^2 &= \langle (xy)^2 \rangle - \langle xy \rangle^2 \\ &= (\langle x \rangle^2 + \sigma_x^2) (\langle y \rangle^2 + \sigma_y^2) - \langle x \rangle^2 \langle y \rangle^2 \\ &= \sigma_x^2 \langle y \rangle^2 + \sigma_y^2 \langle x \rangle^2 + \sigma_x^2 \sigma_y^2 \\ &= \langle x \rangle^2 \langle y \rangle^2 \left[\frac{\sigma_x^2}{\langle x \rangle^2} + \frac{\sigma_y^2}{\langle y \rangle^2} + \frac{\sigma_x^2 \sigma_y^2}{\langle x \rangle^2 \langle y \rangle^2} \right]\end{aligned}\tag{16}$$

となります。標準偏差は

$$\sigma_z = \langle x \rangle \langle y \rangle \sqrt{\frac{\sigma_x^2}{\langle x \rangle^2} + \frac{\sigma_y^2}{\langle y \rangle^2} + \frac{\sigma_x^2 \sigma_y^2}{\langle x \rangle^2 \langle y \rangle^2}}\tag{17}$$

標準偏差と平均との比は

$$\frac{\sigma_z}{\langle z \rangle} = \sqrt{\frac{\sigma_x^2}{\langle x \rangle^2} + \frac{\sigma_y^2}{\langle y \rangle^2} + \frac{\sigma_x^2 \sigma_y^2}{\langle x \rangle^2 \langle y \rangle^2}}\tag{18}$$

という式で表されます。\$|\sigma_x/\langle x \rangle| \ll 1, |\sigma_y/\langle y \rangle| \ll 1\$ のときには、標準偏差はと平均との比は

$$\frac{\sigma_z}{\langle z \rangle} \sim \sqrt{\frac{\sigma_x^2}{\langle x \rangle^2} + \frac{\sigma_y^2}{\langle y \rangle^2}} \quad (19)$$

と書けるので、積の相対誤差は相対誤差の二乗和の平方根で近似できると言えます。

3.3.3 わり算

二つの変数 \$x, y\$ の商 \$z = x/y\$ の平均は、

$$\begin{aligned} \langle z \rangle &= \left\langle \frac{x}{y} \right\rangle \\ &= \langle x \rangle \left\langle \frac{1}{y} \right\rangle \end{aligned} \quad (20)$$

となりますが、\$\left\langle \frac{1}{y} \right\rangle\$ は必ずしも \$\frac{1}{\langle y \rangle}\$ と一致するとは限りません。
\$y\$ の分布が正規分布に従い、分布密度関数が

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{(y - \langle y \rangle)^2}{2\sigma_y^2}\right] \quad (21)$$

にしたがうとすると、

$$\left\langle \frac{1}{y} \right\rangle = \frac{1}{\sqrt{2\pi}\sigma_y} \int_{-\infty}^{\infty} \frac{1}{y} \exp\left[-\frac{(y - \langle y \rangle)^2}{2\sigma_y^2}\right] dy \quad (22)$$

ここで \$(y - \langle y \rangle)^2 / 2\sigma_y^2 \equiv t\$ とおけば、

$$\begin{aligned} y &= \langle y \rangle \pm \sqrt{2}\sigma_y t^{1/2} \\ dy &= \pm 2^{-1/2}\sigma_y t^{-1/2} \end{aligned}$$

であり、

$$\begin{array}{lcl} y: & -\infty & \rightarrow \langle y \rangle \rightarrow \infty \\ t: & \infty & \rightarrow 0 \rightarrow \infty \\ \text{符号:} & & (-) \quad (+) \end{array}$$

から

$$\begin{aligned} \left\langle \frac{1}{y} \right\rangle &= \frac{1}{\sqrt{2\pi}\sigma_y} \left[\int_{\infty}^0 \frac{-2^{-1/2}\sigma_y t^{-1/2}}{\langle y \rangle - \sqrt{2}\sigma_y t^{1/2}} e^{-t} dt + \int_0^{\infty} \frac{2^{-1/2}\sigma_y t^{-1/2}}{\langle y \rangle + \sqrt{2}\sigma_y t^{1/2}} e^{-t} dt \right] \\ &= \frac{1}{2\sqrt{\pi}} \int_0^{\infty} \left(\frac{t^{-1/2}}{\langle y \rangle - \sqrt{2}\sigma_y t^{1/2}} + \frac{t^{-1/2}}{\langle y \rangle + \sqrt{2}\sigma_y t^{1/2}} \right) e^{-t} dt \\ &= \frac{\langle y \rangle}{2\sqrt{\pi}\sigma_y^2} \int_0^{\infty} \frac{t^{-1/2} e^{-t} dt}{\langle y \rangle^2 / 2\sigma_y^2 - t} \\ &= \frac{1}{\sqrt{\pi}\langle y \rangle} \int_0^{\infty} t^{-1/2} \left(1 - \frac{2\sigma_y^2 t}{\langle y \rangle^2} \right)^{-1} e^{-t} dt \end{aligned} \quad (23)$$

となります。ここで \$(1 - x)^{-1} = 1 + x + x^2 + x^3 + \dots\$ の関係から

$$\begin{aligned} \left\langle \frac{1}{y} \right\rangle &= \frac{1}{\sqrt{\pi}\langle y \rangle} \int_0^{\infty} t^{-1/2} \left[1 + \frac{2\sigma_y^2 t}{\langle y \rangle^2} + \left(\frac{2\sigma_y^2 t}{\langle y \rangle^2} \right)^2 t^2 + \dots \right] e^{-t} dt \\ &= \frac{1}{\sqrt{\pi}\langle y \rangle} \int_0^{\infty} \left[t^{-1/2} + \frac{2\sigma_y^2}{\langle y \rangle^2} t^{3/2} + \left(\frac{2\sigma_y^2}{\langle y \rangle^2} \right)^2 t^{7/2} + \dots \right] e^{-t} dt \end{aligned} \quad (24)$$

と書けます。

この式は、以下の式で定義される Γ (ガンマ) 関数という関数を使って、さらに書き直すことができます。

$$\Gamma(\alpha) \equiv \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (25)$$

ガンマ関数は

$$\begin{aligned} \Gamma(1/2) &= \sqrt{\pi} \\ \Gamma(1) &= 1 \\ \Gamma(3/2) &= \sqrt{\pi}/2 \\ \Gamma(2) &= 1 \\ \Gamma(5/2) &= 3\sqrt{\pi}/4 \\ \Gamma(3) &= 2 \\ &\dots \dots \dots \\ \Gamma(\alpha+1) &= \alpha\Gamma(\alpha) \end{aligned} \quad (26)$$

のような性質をもっています。

このことから、

$$\begin{aligned} \left\langle \frac{1}{y} \right\rangle &= \frac{1}{\sqrt{\pi}\langle y \rangle} \left[\Gamma\left(\frac{1}{2}\right) + \Gamma\left(\frac{5}{2}\right) \frac{2\sigma_y^2}{\langle y \rangle^2} + \Gamma\left(\frac{9}{2}\right) \left(\frac{2\sigma_y^2}{\langle y \rangle^2}\right)^2 + \dots \right] \\ &= \frac{1}{\langle y \rangle} \left[1 + \frac{3}{2} \cdot \frac{1}{2} \left(\frac{2\sigma_y^2}{\langle y \rangle^2}\right) + \frac{7}{2} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \left(\frac{2\sigma_y^2}{\langle y \rangle^2}\right)^2 + \dots \right] \\ &= \frac{1}{\langle y \rangle} \left[1 + \frac{3}{2} \left(\frac{\sigma_y}{\langle y \rangle}\right)^2 + \frac{105}{4} \left(\frac{\sigma_y}{\langle y \rangle}\right)^4 + \dots \right] \\ &\sim \frac{1}{\langle y \rangle} \end{aligned} \quad (27)$$

また、商の分散は

$$\begin{aligned} \left\langle \left(\frac{x}{y} - \left\langle \frac{x}{y} \right\rangle \right)^2 \right\rangle &= \langle x^2 \rangle \left\langle \frac{1}{y^2} \right\rangle - \langle x \rangle^2 \left\langle \frac{1}{y} \right\rangle^2 \\ &= (\langle x \rangle^2 + \sigma_x^2) \left\langle \frac{1}{y^2} \right\rangle - \langle x \rangle^2 \left\langle \frac{1}{y} \right\rangle^2 \end{aligned} \quad (28)$$

$$\begin{aligned} \left\langle \frac{1}{y^2} \right\rangle &= \frac{1}{\sqrt{2\pi}\sigma_y} \int_{-\infty}^{\infty} \frac{1}{y^2} \exp\left[-\frac{(y-\langle y \rangle)^2}{2\sigma_y^2}\right] dy \\ &= \frac{1}{\sqrt{2\pi}\sigma_y} \left[\int_{-\infty}^0 \frac{-2^{-1/2}\sigma_y t^{-1/2}}{(\langle y \rangle - \sqrt{2}\sigma_y t^{1/2})^2} e^{-t} dt + \int_0^{\infty} \frac{2^{-1/2}\sigma_y t^{-1/2}}{(\langle y \rangle + \sqrt{2}\sigma_y t^{1/2})^2} e^{-t} dt \right] \\ &= \frac{1}{2\sqrt{\pi}} \int_0^{\infty} \left[\frac{t^{-1/2}}{(\langle y \rangle - \sqrt{2}\sigma_y t^{1/2})^2} + \frac{t^{-1/2}}{(\langle y \rangle + \sqrt{2}\sigma_y t^{1/2})^2} \right] e^{-t} dt \\ &= \frac{1}{2\sqrt{\pi}} \int_0^{\infty} \frac{2t^{-1/2}(\langle y \rangle^2 + 2\sigma_y^2 t)}{(\langle y \rangle^2 - 2\sigma_y^2 t)^2} e^{-t} dt \\ &= \frac{1}{\sqrt{\pi}\langle y \rangle^2} \int_0^{\infty} \frac{t^{-1/2}(1 + 2\sigma_y^2 t/\langle y \rangle^2)}{(1 - 2\sigma_y^2 t/\langle y \rangle^2)^2} e^{-t} dt \end{aligned} \quad (29)$$

となります。ここでは $(1+x)/(1-x)^{-2} = 1 + 3x + 5x^2 + 7x^3 + \dots$ の関係から

$$\begin{aligned}
 \left\langle \frac{1}{y^2} \right\rangle &= \frac{1}{\sqrt{\pi}\langle y \rangle^2} \int_0^\infty t^{-1/2} \left[1 + 3 \frac{2\sigma_y^2 t}{\langle y \rangle^2} + 5 \left(\frac{2\sigma_y^2 t}{\langle y \rangle^2} \right)^2 t^2 + \dots \right] e^{-t} dt \\
 &= \frac{1}{\sqrt{\pi}\langle y \rangle^2} \int_0^\infty \left[t^{-1/2} + 3 \cdot \frac{2\sigma_y^2}{\langle y \rangle^2} t^{3/2} + 5 \left(\frac{2\sigma_y^2}{\langle y \rangle^2} \right)^2 t^{7/2} + \dots \right] e^{-t} dt \\
 &= \frac{1}{\sqrt{\pi}\langle y \rangle^2} \left[\Gamma\left(\frac{1}{2}\right) + 3 \cdot \frac{2\sigma_y^2}{\langle y \rangle^2} \Gamma\left(\frac{5}{2}\right) + 5 \left(\frac{2\sigma_y^2}{\langle y \rangle^2} \right)^2 \Gamma\left(\frac{9}{2}\right) + \dots \right] \\
 &= \frac{1}{\langle y \rangle^2} \left[1 + 3 \cdot \frac{3}{2} \cdot \frac{1}{2} \frac{2\sigma_y^2}{\langle y \rangle^2} + 5 \cdot \frac{7}{2} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \left(\frac{2\sigma_y^2}{\langle y \rangle^2} \right)^2 + \dots \right] \\
 &= \frac{1}{\langle y \rangle^2} \left[1 + \frac{9}{2} \left(\frac{\sigma_y}{\langle y \rangle} \right)^2 + \frac{525}{4} \left(\frac{\sigma_y}{\langle y \rangle} \right)^4 + \dots \right] \tag{30}
 \end{aligned}$$

と書けます。また、

$$\left\langle \frac{1}{y} \right\rangle^2 = \frac{1}{\langle y \rangle^2} \left[1 + 3 \left(\frac{\sigma_y}{\langle y \rangle} \right)^2 + \frac{219}{4} \left(\frac{\sigma_y}{\langle y \rangle} \right)^4 + \dots \right] \tag{31}$$

ですから、式 (28) から

$$\begin{aligned}
 \left\langle \left(\frac{x}{y} - \left\langle \frac{x}{y} \right\rangle \right)^2 \right\rangle &= (\langle x \rangle^2 + \sigma_x^2) \left\langle \frac{1}{y^2} \right\rangle - \langle x \rangle^2 \left\langle \frac{1}{y} \right\rangle^2 \\
 &= \frac{\langle x \rangle^2 + \sigma_x^2}{\langle y \rangle^2} \left[1 + \frac{9}{2} \left(\frac{\sigma_y}{\langle y \rangle} \right)^2 + \frac{525}{4} \left(\frac{\sigma_y}{\langle y \rangle} \right)^4 + \dots \right] \\
 &\quad - \frac{\langle x \rangle^2}{\langle y \rangle^2} \left[1 + 3 \left(\frac{\sigma_y}{\langle y \rangle} \right)^2 + \frac{219}{4} \left(\frac{\sigma_y}{\langle y \rangle} \right)^4 + \dots \right] \tag{32}
 \end{aligned}$$

3.4 畳み込み

変数 x, y がそれぞれ確率密度関数 $f(x), g(y)$ で表されるような統計分布に従う場合、 $z = x + y$ はどのような統計分布に従うでしょうか？