誤差の伝播

- いきなり目的の量が計測できない場合
  ⇒　機構による変換を行って、別の物理量にして計測
- 和、差、積、商、定数倍、累乗、平方根、任意の関数
- その場合、計測誤差がわかっているのは変換後の物理量
  ⇒元の目的の量の誤差はどう見積もるのか？

和、差、積、商による誤差の伝播則

- 和の場合

$$q = (x_{best} \pm \delta x) + (y_{best} \pm \delta y)$$
$$= (x_{best} + y_{best}) \pm (\delta x + \delta y)$$

よって

$$q_{best} = x_{best} + y_{best}$$
$$\delta q = \delta x + \delta y$$

つまり差と同じ

和と差の誤差伝播則

$$q = x + \cdots + z - (u + \cdots + w)$$

の絶対誤差はそれぞれの絶対誤差の和

$$\delta q = \delta x + \cdots + \delta z + \delta u + \cdots \delta w$$

ここでチェック：質量の和と差

- 二つのフラスコ　空のときと液体入りの場合の質量
  M_1 = 540±10 g (液入り)　m_1 = 72±1 g (空)
  M_2 = 940±20 g (液入り)　m_2 = 97±1 g (空)

液体の全質量は？
M_{best} = (540 - 72 + 940 - 97)g = 1311 g
δM = (10 + 1 + 20 + 1) g = 32 g
誤差の有効桁は？　M_1,m_1,M_2,m_2ともに一桁！
よってまるめて　δM=30 g
最良推定値の有効数字の最終桁をあわせて、
M = 1310±30 g

⇒　フラスコの質量の誤差が液体込みの誤差より1桁小さいので、全体の誤差に影響していないことがわかる。

積、商の場合

- **<u>積は相対誤差の和</u>　⇒　商の場合は？**

$$x = x_{best}\left(1 \pm \frac{\delta x}{|x_{best}|}\right) \qquad y = y_{best}\left(1 \pm \frac{\delta y}{|y_{best}|}\right)$$

$$q = \frac{x}{y} = \frac{x_{best}}{y_{best}}\left(1 \pm \frac{\delta x}{|x|}\right) \Big/ \left(1 \pm \frac{\delta y}{|y|}\right)$$

- q**最大なのは** $\frac{x_{best}}{y_{best}}\left(1 + \frac{\delta x}{|x|}\right) \Big/ \left(1 - \frac{\delta y}{|y|}\right)$ **のとき**
- q**最小なのは** $\frac{x_{best}}{y_{best}}\left(1 - \frac{\delta x}{|x|}\right) \Big/ \left(1 + \frac{\delta y}{|y|}\right)$ **のとき**

積、商の場合

- **ここで近似計算** dx**が1より十分小さければ**（<u>dx<<1と書く</u>）

$$\frac{1}{1+\delta x} \approx 1 - \delta x \qquad \frac{1}{1-\delta x} \approx 1 + \delta x$$

といえる。

- **なぜか？**
  - **二項定理（問題3．8）によるもの**
  - **あるいはテーラー展開近似から** $f(x_0 + \delta x) \approx f(x_0) + f'(x_0)\delta x$

f(x)= 1/x**と考えると、**

$$f(x_0 + \delta x) = 1/(x_0 + \delta x) \approx 1/x_0 - 1/x_0{}^2 \delta x$$

x_0=1**を代入すると、** $\frac{1}{1+\delta x} \approx 1 - \delta x$

よく使う近似式

$$(1 \pm \delta x)^n \approx 1 \pm n\delta x \quad (\delta x << 1)$$

$$\sin(\delta x) \approx \delta x \quad (\delta x << 1)$$

いずれもテーラー展開から説明される。

$$f(x+\delta x) = f(x) + \frac{f^1(x)}{1!}\delta x + \frac{f^2(x)}{2!}\delta x^2 + \cdots + \frac{f^n(x)}{n!}\delta x^n + \cdots$$

$f^n(x)$　f**の**n**階微分**

n!　n**の階乗**　n×(n-1)×(n-2)×…×2×1

よく使う近似式

$\frac{1}{1-\delta x} \approx 1 + \delta x$　**をつかうと、以下の近似が成り立つ**

$$\frac{1+\delta y}{1-\delta x} \approx (1+\delta x)(1+\delta y) = 1 + \delta x + \delta y + \underline{\delta x \delta y}$$

$$\approx 1 + \delta x + \delta y$$

δx、δy<<1のとき
高次項は省略できる

商の誤差の近似計算に使える

積、商の場合

$$\frac{1+\delta y}{1-\delta x} \approx 1+\delta x+\delta y$$ を使うと、商の最大最小は？

$$q_{\max} = \frac{x_{best}}{y_{best}}\left(1+\frac{\delta x}{|x|}\right)\Big/\left(1-\frac{\delta y}{|y|}\right) \approx \frac{x_{best}}{y_{best}}\left(1+\frac{\delta x}{|x|}+\frac{\delta y}{|y|}\right)$$

$$q_{\min} = \frac{x_{best}}{y_{best}}\left(1-\frac{\delta x}{|x|}\right)\Big/\left(1+\frac{\delta y}{|y|}\right) \approx \frac{x_{best}}{y_{best}}\left(1-\frac{\delta x}{|x|}-\frac{\delta y}{|y|}\right)$$

つまり商の最良推定値と誤差は

$$q_{best} = \frac{x_{best}}{y_{best}}$$

⇒積と同じになる

$$\frac{\delta q}{|q_{best}|} = \frac{\delta x}{|x_{best}|} + \frac{\delta y}{|y_{best}|}$$

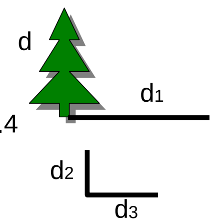
積と商の誤差伝播則

$$q = \frac{x \times \cdots \times z}{u \times \cdots \times w}$$

の相対誤差はそれぞれの相対誤差の和

$$\frac{\delta q}{|q|} = \frac{\delta x}{|x|} + \cdots \frac{\delta z}{|z|} + \frac{\delta u}{|u|} + \cdots \frac{\delta w}{|w|}$$

ここでチェック：高い木の高さの測量



- 適当な棒の影の長さをつかって比例計算
  $d = d_1 d_2 / d_3$
  $d_1 = 200 \pm 2$, $d_2 = 5.5 \pm 0.1$, $d_3 = 10.0 \pm 0.4$

  $d_{best} = 200 \times 5.5/10.0 = 110$ (ft:フィート)
- 相対誤差
  $\delta d/d = 2/200 + 0.1/5.5 + 0.4/10.0 = (1+2+4)\% = 7\%$
- 絶対誤差
  $\delta d = 110 \times 7\% = 7.7$　　8(ft)

  したがって、木の高さは　$l = 110 \pm 8$ ft

特殊な場合：定数倍

- q=Bxの誤差は？　Bは既知の定数で誤差がないとする。
- 積の場合の誤差伝播だから、

  $\delta q/|q| = \delta B/|B| + \delta x/|x| = \delta x/|x|$　(xの相対誤差と同じ)

  よって　$\delta q = |Bx|\,\delta x/|x| = |B|\,\delta x$
- 小さい値を計測したいときは、何回分かを積み重ねて計測すると誤差を小さくすることができる。
- 例：　紙1枚の厚さ
  200枚重ねて測る　⇒ T = 1.3±0.1 (in:インチ)
  一枚分だと
  $t = 1/200\ T = 0.0065 \pm 0.0005$(in) (= $(6.5 \pm 0.5) \times 10^{-3}$in)
  すべての紙の厚さが同じとわかっている場合のみ。

特殊な場合：べき乗

- $v^2 = v \times v$だから、
  $\delta(v^2)/v = 2 \times \delta v/|v|$　となる。

- 一般に　$q = x^n$ の相対誤差は
  $$\delta q/|q| = n \times \delta x/|x|$$
  一般にはnは実数でよい（関数による誤差伝播で扱う）

- 例：立方体の体積

誤差解析の初歩

- いくつかの計測値から知りたい数値を計算する場合、どの計測値の誤差を少なくすると最終的な誤差が少なくできるか。

- 例：重力加速度gの測定
  高さhから石が落下する速度を測定
  $t = 1.6 \pm 0.1$s
  $h = 46.2 \pm 0.3$ft

  $h = 1/2\, gt^2$　より $g = 2h/t^2 = 2 \times 46.2/1.6^2 = 36.1$ ft

  このgの値の誤差はどのくらいと見積もるべきか？

誤差解析の初歩

$t = 1.6 \pm 0.1$s　　$h = 46.2 \pm 0.3$ft

$h = 1/2\, gt^2$　より $g = 2h/t^2 = 2 \times 46.2/1.6^2 = 36.1$ ft

相対誤差　$\delta g/|g| = \delta h/|h| + 2 \times \delta t/|t|$
$= 0.3/46.2 + 2 \times 0.1/1.6$
$= 13.3\%$

よって
$\delta g = 0.133 \times 36.1 = 4.80$　　5 ft/s²

ゆえに
$g = 36 \pm 5$ ft/s²
（最良推定量の有効桁を誤差の最終桁とあわせる）

誤差解析の初歩

相対誤差　$\delta g/|g| = \delta h/|h| + 2 \times \delta t/|t|$
$= 0.3/46.2 + 2 \times 0.1/1.6$
$= \underline{0.7\%} + \underline{2 \times 6.3\%}$
$=13.3\%$

tの誤差のほうが18倍も大きい

いくらhの誤差を小さくしても6%以上の誤差はtの誤差が原因で残ってしまう
⇒hよりもtの誤差を小さくするほうが賢明

独立な誤差

- 二つの測定値の和の誤差 ⇒ 絶対誤差の和
- しかし二つの誤差は互いに独立にランダムであることがほとんど。

⇒ 誤差が絶対誤差の和になることは、どちらかが最大、どちらかが最小の場合のみであり、めったにおきない。

⇒絶対誤差の和では、過大評価の可能性がある。

独立な誤差

- 誤差の範囲の意味
たとえば、95%でその範囲にはいるということ
(100%の範囲を示すのは現実的ではない)
- 和の誤差の95%範囲は誤差の和よりは小さいことが統計的に示される
  - ともに独立な正規分布の場合は、和の誤差は絶対誤差の二乗和になる
- 二乗和: $\delta q = \sqrt{\delta x^2 + \delta y^2}$

誤差の単純和と二乗和の使い分け

- 2つの誤差に関係(相関という)があれば
例:巻尺が伸び縮みしている場合

このときにはxが過大に測られた場合は、yも過大、xが過小ならyも過小に測られる。この場合には和の誤差は二乗和ではなく絶対誤差の和となる

ここでチェック: 2つのビーカーにはいった水の体積

- V1 = 130±6ml, V2 = 65±4ml
両方の誤差が独立でランダムだとすると、両者を足し合わせ水の体積は？
- 誤差は二乗和で与えられる
⇒ $\delta V = \sqrt{6^2+4^2} = 7$ml ⇒ 195±7ml
- もし誤差が独立でないとしたら？
⇒ $\delta V = 6+4 = 10$ml ⇒ 195±10ml

和と差、積と商の誤差伝播則
二乗和バージョン

x,…z,u,…wの誤差が**ランダムで独立**の場合

$$q = x + \cdots + z - (u + \cdots + w)$$

の絶対誤差はそれぞれの絶対誤差の二乗和で与えられる

$$\delta q = \sqrt{\delta x^2 + \cdots + \delta z^2 + \delta u^2 + \cdots \delta w^2}$$

$$q = \frac{x \times \cdots \times z}{u \times \cdots \times w}$$

の相対誤差はそれぞれの相対誤差の和

$$\frac{\delta q}{|q|} = \sqrt{\left(\frac{\delta x}{|x|}\right)^2 + \cdots \left(\frac{\delta z}{|z|}\right)^2 + \left(\frac{\delta u}{|u|}\right)^2 + \cdots \left(\frac{\delta w}{|w|}\right)^2}$$

チェック2：直流モーターの効率

- 質量mのおもりをモーターで高さhまで持ち上げる

おこなった仕事はmgh、モーターにかけた電圧V，電流I，モータの作動時間をtとすると，消費された電気エネルギーはVIt

よって効率　e　=　mgh/VIt

m,h,V,Iは1%の誤差，tは5%の誤差で計測できたとする。
もし，誤差がランダムで独立なら

$$\frac{\delta e}{e} = \sqrt{\left(\frac{\delta m}{m}\right)^2 + \left(\frac{\delta h}{h}\right)^2 + \left(\frac{\delta V}{V}\right)^2 + \left(\frac{\delta I}{I}\right)^2 + \left(\frac{\delta t}{t}\right)^2}$$

$$= \sqrt{1+1+1+1+25}\% = \sqrt{29}\% \approx 5\%$$

5倍程度の相対誤差の差があると小さい誤差は無視しても結果に影響しない，つまり大きな誤差が支配的であることがわかる．

任意の一変数関数による誤差伝播

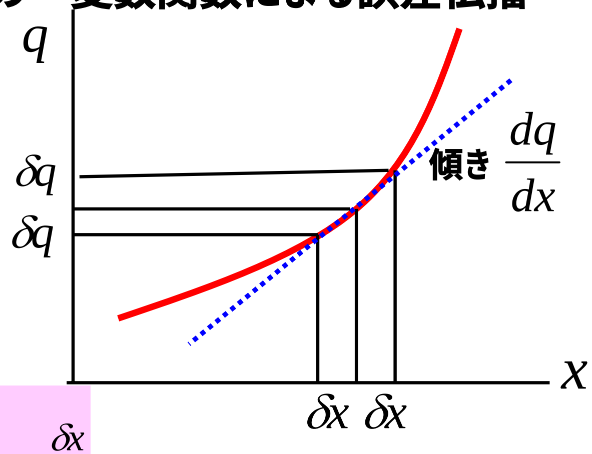
- ガラスの屈折率nと臨界角θの関係

臨界角：入射光の角度を浅くしていくと全反射して中に入り込まくなる、その角度

n=1/sin **θ**　←　和積では書けない

θの誤差δθがnの誤差δnに及ぼす影響はどのくらいか？

任意の一変数関数による誤差伝播



$$\delta q = \left|\frac{dq}{dx}\right|_{x=x_{best}} \delta x$$

任意の一変数関数における誤差

x=x$_{best}$±δxのとき、任意の一変数関数q=q(x)の誤差は以下のようになる

$$q_{best} = q(x_{best})$$

$$\delta q = \left|\frac{dq}{dx}\right|_{x=x_{best}} \delta x$$

任意の一変数関数における誤差

- 例:
  θ＝20±3degのとき, n=1/sinθの誤差は？

$$\delta n = \left| -\frac{\cos\theta}{\sin^2\theta} \right| \delta\theta$$

sin(20deg) = 0.34, cos(20deg) = 0.94
δθ= 3deg = 0.05rad
⇒ラジアンで書かないといけない

ゆえに
δn = 0.94/0.34^2×0.05 = 0.41

特別な場合 ： 指数関数

- q = x^nの場合
  δq = |nx^{n-1}|δx
  両辺を|q|=|x^n|で割ると
  δq/|q| = |n|δx/|x|

つまり, 指数関数の相対誤差はもとの相対誤差の|n|倍.
ただし、nは実数でよい。

- q = √x ⇒ $x^{1/2}$
  δq/|q| = 1/2 δx/|x|

また, q=1/xの相対誤差はxの相対誤差と同じである.

ここでチェック：指数関数の誤差伝播

- xの測定値が100±6のとき, √xは誤差を含めてどのような値となるか

δq/|q| = 1/2 δx/|x|
xの相対誤差δx/|x| = 6/100 = 6%
δq/|q| = 3%
√xの最良推定値：√100 = 10.0 (有効数字3桁)
δq = 10x0.03 = 0.3(有効数字1桁)
ゆえに 10.0±0.3

誤差の逐次伝播

- より複雑な計算式の場合，誤差の伝播をどう計算するか．
  ⇒以下のステップに分解できるはず
  1) 和もしくは差
  2) 積もしくは商
  3) x^n, sin(x),exp(x), ln(x)などの一変数関数

  各ステップごとに伝播する誤差を逐次的に計算していけばよい．

- 例：$q = x(y - z\sin u)$

$$\delta(\sin u) = \delta u\,|\cos u|$$

$$\frac{\delta(z\sin u)}{|z\sin u|} = \frac{\delta z}{|z|} + \frac{\delta(\sin u)}{|\sin u|} = \frac{\delta z}{|z|} + \delta u\left|\frac{\cos u}{\sin u}\right|$$

$$\delta(y - z\sin u) = \delta y + \delta(z\sin u)$$

$$= \delta y + \left(\frac{\delta z}{|z|} + \delta u\left|\frac{\cos u}{\sin u}\right|\right)|z\sin u|$$

$$\frac{\delta(x(y - z\sin u))}{|x(y - z\sin u)|} = \frac{\delta x}{|x|} + \frac{\delta(y - z\sin u)}{|y - z\sin u|}$$

$$= \frac{\delta x}{|x|} + \frac{\delta y + \left(\frac{\delta z}{|z|} + \delta u\left|\frac{\cos u}{\sin u}\right|\right)|z\sin u|}{|y - z\sin u|}$$

ここでチェック：逐次伝播

- x=200±2, y=50±2, z=40±2
  誤差が独立かつランダムだった場合，q=x/(y-z)は？

  y-zの最良推定値：50-40 =10
  y-zの誤差の2乗：$(\delta(y-z))^2 = 2^2+2^2 = 8$
  y-zの相対誤差の2乗: $(\delta(y-z)/|y-z|)^2 = 8/10^2 = 0.08$
  xの相対誤差の2乗：$(\delta x/|x|)^2 = (2/200)^2 = 0.0001$
  x/(y-z)の相対誤差：
  $\delta(x/(y-z))/|x/(y-z)| = \sqrt{(\delta x/|x|)^2 + (\delta(y-z)/|y-z|)^2}$
  $= \sqrt{0.08 + 0.0001} = \sqrt{0.08}$
  x/(y-z)の最良推定値：200/10 = 20 (有効数字2桁)
  x/(y-z)の絶対誤差：$20\times\sqrt{0.08} = 6$
  ゆえに x/(y-z) = 20±6

複雑な計算式による誤差伝播

- 逐次的に計算すると誤差を過小に見積もってしまう場合がある。
- 例：xy-xz, x+y/x+z
  xyとxzを比べると、xを大きく評価した場合はxyもxzともに大きくなり、xを小さく評価した場合はともに小さくなる
  ⇒ したがってxyとxzの差はそれほど大きく変化しない
  ⇒ これを誤差の打ち消しまたは誤差の補償という。

- 同じ変数が2回以上現れる場合にこのようなことが起こることがある
  その場合逐次計算してはいけない。
  ⇒ どうするか？

複雑な計算式による誤差伝播

1) **同じ変数が2回以上現れないように変形する**

xy-xz = x(y-z)

⇒ y-zの誤差を計算してから、xとの積の誤差を計算する

注：必ずしも変形できるとはかぎらない

2) **微分演算をつかって一回で誤差の計算を行う**

微分演算による誤差伝播の計算

- **一変数関数q(x)の場合、qの誤差はxの誤差を用いて**

  δq=|dq/dx|**δ**x

  とかけた。

- **二変数関数q(x,y)の場合は？**

微分演算による誤差伝播の計算

$$q(x+\delta x, y+\delta y) \approx q(x,y) + \frac{\partial q}{\partial x}\delta x + \frac{\partial q}{\partial y}\delta y$$

$$q(x-\delta x, y-\delta y) \approx q(x,y) - \frac{\partial q}{\partial x}\delta x - \frac{\partial q}{\partial y}\delta y$$

⇒ 偏微分の連鎖定理を用いた近似

よって、

$$\delta q = \left|\frac{\partial q}{\partial x}\right|\delta x + \left|\frac{\partial q}{\partial y}\right|\delta y$$

⇒一般にn変数関数の場合にも同様にいえる。

多変数関数における誤差

x,...,z,の測定誤差をδx,..., δzとし、これらの誤差が独立でランダムなとき、関数q(x,...,z)の誤差は

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \cdots + \left(\frac{\partial q}{\partial z}\delta z\right)^2}$$

となる。

この値は単純和より大きくなることはない。

$$\delta q \le \left|\frac{\partial q}{\partial x}\right|\delta x + \cdots + \left|\frac{\partial q}{\partial z}\right|\delta z$$

多変数関数における誤差

- 例: $q(x,y)=x+y$

$$\frac{\partial q}{\partial x}=\frac{\partial q}{\partial y}=1$$

$$\therefore \delta q=\sqrt{\delta x^2+\delta y^2}$$

$q(x,y)=xy$

$$\frac{\partial q}{\partial x}=y,\quad \frac{\partial q}{\partial y}=x$$

$$\therefore \delta q=\sqrt{(y\delta x)^2+(x\delta y)^2}$$

$$\therefore \frac{\delta q}{|q|}=\frac{\sqrt{(y\delta x)^2+(x\delta y)^2}}{|xy|}$$

$$=\sqrt{\left(\frac{\delta x}{|x|}\right)^2+\left(\frac{\delta y}{|y|}\right)^2}$$

⇒たしかに以前の法則がそのまま成立している

ここでチェック:

- $q=x^2y-xy^2$の値を求めるため、xとyを計測したところ、

x= 3.0±0.1　　y= 2.0±0.1

となった。qの最良推定値と誤差はどう見積もればよいか。ただし、x,yの誤差はランダムで独立とする。

q_{best} = 3.0^2×2.0−3.0×2.0^2 = 6.0

∂q/∂x= 2xy−y^2 = 8.0,　∂q/∂y= x^2 −2xy = −3.0

δq = √(8.0×0.1)^2 ＋(3.0×0.1)^2

　= 0.85.. = 0.9

よって　q = 6.0±0.9

ランダムな誤差の統計的取り扱い

- 誤差には2種類ある(既に述べた)
  - ランダム誤差 (random uncertainty)
  - 系統誤差 (systematic uncertainty)
- ランダムな誤差は統計的に扱うことができる
  - 平均値
  - 標準偏差、分散
  - 確率分布(正規分布、二項分布、他)

平均値は誤差が正規分布するときに重要な意味を持つ

平均値、標準偏差、分散の定義

- 平均値 (average, mean)

$$\bar{x}=\frac{x_1+x_2+\cdots+x_N}{N}=\sum_{i=1}^{N}x_i=\sum x_i$$

- 分散 (variance)

$$S=\frac{1}{N}\sum\underline{(x_i-\bar{x})^2}$$

平均値と各データの差(残差:residual)

- 標準偏差 (standard deviation: SD)

$$\sigma=\sqrt{S}=\sqrt{\frac{1}{N}\sum(x_i-\bar{x})^2}$$

平均と分散の意味

- **平均値**
  - **最良推定値として使われる (代表値)**
- **分散、標準偏差**
  - **平均値の周りでどのくらいデータがばらついているかを示す**

    ⇒ ランダム誤差の程度を表すといえる

    ⇒ 何回も測定をすると、分散からランダム誤差の大きさを見積もることができる

ここでチェック：平均、分散、標準偏差

- **次のデータの平均、分散、標準偏差を求めよ**

71, 72, 72, 73, 71

平均： 70+(1+2+2+3+1)／5 = 71.8
分散： $(0.8^2+0.2^2+0.2^2+1.2^2+0.8^2)/5 = 0.56$
標準偏差：$\sqrt{0.56} = 0.75$

ヒストグラム

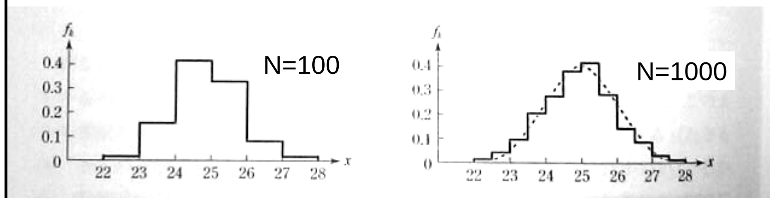


図 5・3 図5・2と同様のヒストグラム，ただし測定データは100個

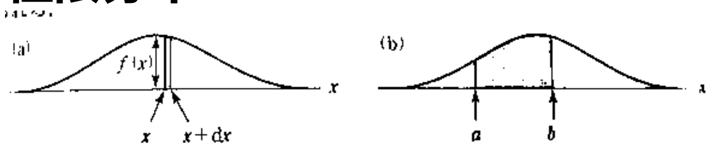
図 5・4 図5・2と同様のヒストグラム，ただし測定データは1000個．破線は極限分布

縦軸をNで割った割合にする ⇒ 相対ヒストグラム

計測数Nをどんどん増やして、区間の幅 (ヒストグラムのビン) を細かくしていくと、連続的な分布に近づく

⇒ 極限分布

極限分布



- f(x) : **確率密度**

  x**と**x+dx**の間をとる確率が**f(x)dx

  a**と**b**の間を取る確率は積分で求められる** $\int_a^b f(x)dx$

  全体で確率は1 $\int_{-\infty}^{\infty} f(x)dx = 1$

極限分布の平均、分散

- 平均 $\bar{x} = \int_{-\infty}^{\infty} x f(x) dx$
- 分散 $\sigma^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 f(x) dx$

$$\sigma^2 = \int_{-\infty}^{\infty} (x^2 - 2\bar{x}x + \bar{x}^2) f(x) dx$$

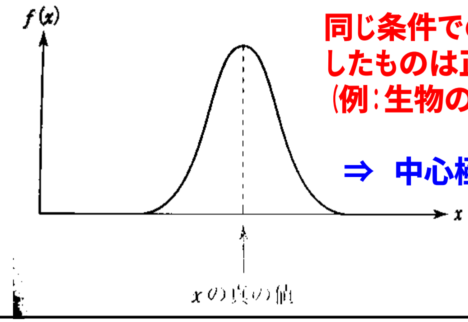
$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \bar{x}^2$$

分散は
二乗平均ー平均の二乗

正規分布 (ガウス分布)

- 極限分布のひとつ

二項分布 (確率pで起こる事象がN回の試行で何回おこったか) のN⇒∞の極限



同じ条件での試行結果を足し算したものは正規分布する
(例：生物の体長、試験の点数)

⇒ 中心極限定理 (詳細は略)

正規分布

平均 $\bar{x}$ 分散 σ^2 の正規分布

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right)$$

f(x)を積分して1にするための係数
(正規化係数)

$\exp(x) \equiv e^x$ 指数関数

信頼限界

- 正規分布に従うとき、平均±σの間に入る確率

$$P(|x-\bar{x}| \le \sigma) = \int_{\bar{x}-\sigma}^{\bar{x}+\sigma} f(x) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{\bar{x}-\sigma}^{\bar{x}+\sigma} \exp(-(x-\bar{x})^2 / 2\sigma^2) dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \exp(-z^2/2) dz \quad \Leftarrow \quad \frac{x-\bar{x}}{\sigma} = z$$

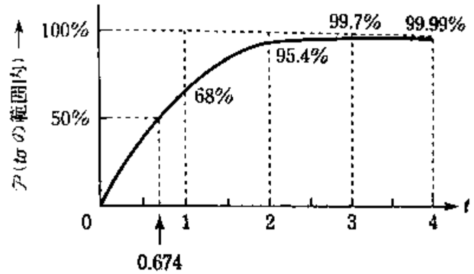
$$\approx 0.68$$

同様に平均±2σの間の場合は、0.95, 3σの間なら、0.99

誤差関数(正規誤差積分)

- 正規分布で、平均±tσの間に入る確率

$$\mathrm{erf}(t) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{t}\exp\left(-\frac{z^2}{2}\right)dz$$



誤差関数(正規誤差積分)

- 正規分布で、平均±tσの間に入る確率

$$\mathrm{erf}(t) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{t}\exp\left(-\frac{z^2}{2}\right)dz$$

⇒ tの関数になるが、解析的には表せないので、数値表の形で提供される

⇒ **教科書巻末付録A**

⇒ C言語のmath関数にもerf(x)がある。

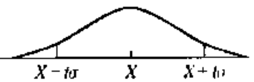
正規誤差積分表の見方

t＝0.24 (平均±0.24σに入る確率)

付録A 正規誤差積分 I

表 A・1 P($t\sigma$ の範囲内) $= \int_{X-t\sigma}^{X+t\sigma} G_{X,\sigma}(x)\,dx$

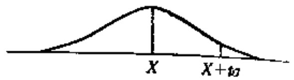
Pをtの関数としてパーセントで表示



t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00	0.80	1.60	2.39	3.19	3.99	4.78	5.58	6.38	7.17
0.1	7.97	8.76	9.55	10.34	11.13	11.92	12.71	13.50	14.28	15.07
0.2	15.85	16.63	17.41	18.19	18.97	19.74	20.51	21.28	22.05	22.82
0.3	23.58	24.34	25.10	25.86	26.61	27.37	28.12	28.86	29.61	30.35
0.4	31.08	31.82	32.55	33.28	34.01	34.73	35.45	36.16	36.88	37.59

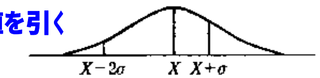
正規誤差積分表の見方

- 平均との差が+tσにはいる確率(さっきの値の半分)
  ⇒ **付録B 正規誤差積分表II**



$t_1 \le t \le t_2$ にはいる確率を求める場合

t1,t2ともに正の場合：
t2に対する表の値からt1に対する値を引く



t1が負、t2が正の場合：
t2に対する表の値とt1に対する値を足す

ここでチェック：正規誤差積分表

- 平均が10.0、標準偏差が2.0の正規分布において、測定値が９と12の間に入る確率を求めよ

測定値が９と１２の間

⇒測定値は丸められているので、8.5から12.5の間を取ると考える

t1 = (8.5-10.0)／2.0 = -0.75

t2 = (12.5-10.0)／2.0 = 1.25

$$\frac{x-\overline{x}}{\sigma}=t$$

付録２の表で、t1に対する値は27.34、t2に対する値は39.44

したがって、両者を足せば 66.7% (答)

平均値の重要性

- 正規分布を仮定すると、測定値の平均値は真の値の最良推定値となる ⇒なぜか？

標本(sample) $x_1,\ldots.,x_N$が測定できたとする。

真の値をX、真の標準偏差をσと仮定する。

このとき、i番目の測定値がx_iとなる確率は

$$P_{X,\sigma}(x_i \le x \le x_i+dx)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x_i-X)^2}{2\sigma^2}\right)dx$$

$$\propto \frac{1}{\sigma}\exp\left(-\frac{(x_i-X)^2}{2\sigma^2}\right)$$

平均値の重要性

- n回の測定がすべて独立とすれば、x1,…xnが測定される確率はxiが得られる確率の積となるので、

$$P_{X,\sigma}(x_1,\cdots x_N)\propto\frac{1}{\sigma}\exp\left(-\frac{(x_1-X)^2}{2\sigma^2}\right)\times\cdots\times\frac{1}{\sigma}\exp\left(-\frac{(x_N-X)^2}{2\sigma^2}\right)$$

$$=\frac{1}{\sigma^N}\exp\left(-\frac{\sum_i (x_i-X)^2}{2\sigma^2}\right)$$

この確率Pを最大化する、つまり最も測定値がおこりやすいXは？ ⇒ XでPを偏微分して０とおけばよい。

$$\sum_{i=1}^{n}(x_i-X)=\sum_{i=1}^{n}x_i-NX=0 \qquad \boxed{\therefore X=\frac{\sum x_i}{N}}$$

平均値の重要性

同一のものに対する複数の計測値が正規分布に従う場合、標本平均値

$$\frac{\sum x_i}{N}$$

はもっとも確からしい真の値の推定値を与える（最尤推定値）。

標準偏差の最良推定値

- 真の値Xの最良推定値は平均で与えられた
- Xは計測して初めて値が推定できるので、その計測方法によってきまる**計測自体の不確かさを見積もる**ことは重要
  ⇒ いきなり真の値Xの分布ではなく、
  まず計測誤差の分布を見積もりたい
  (真の値Xの推定値 (平均) の分布と誤差はあとで導出する)

標準偏差の最良推定値

$$P_{X,\sigma}(x_1,\cdots x_n) \propto \frac{1}{\sigma^N}\exp\left(-\frac{\sum_i (x_i - X)^2}{2\sigma^2}\right)$$

この確率Pを最大化する、つまり**最も測定値がおこりやすいσは？** ⇒ 以下のQを最小化すればよい
⇒σでQを偏微分して0とおけばよい。

$$Q = -\log P_{X,\sigma}(x_1,\cdots x_n) = N\log\sigma + \frac{\sum_i (x_i - X)^2}{2\sigma^2}$$

標準偏差の最良推定値

$$Q = -\log P_{X,\sigma}(x_1,\cdots x_n) = N\log\sigma + \frac{\sum_i (x_i - X)^2}{2\sigma^2}$$

$$\frac{\partial Q}{\partial \sigma} = \frac{N}{\sigma} - \frac{1}{\sigma^3}\sum_i (x_i - X)^2 = 0$$

$$\therefore \sigma^2 = \frac{1}{N}\sum_i (x_i - \underline{X})^2$$

真の値Xは未知である。
最良推定値は標本平均 $\overline{X}$

標準偏差の最良推定値

$$\sigma^2 = \frac{1}{N}\sum_i (x_i - X)^2 = \frac{1}{N}\sum_i (x_i - \bar{x} + \bar{x} - X)^2$$

$$= \frac{1}{N}\sum_i \left\{(x_i - \bar{x})^2 + (\bar{x} - X)^2 + (x_i - \bar{x})(\bar{x} - X)\right\}$$

$$= \frac{1}{N}\sum_i (x_i - \bar{x})^2 + (\bar{x} - X)^2 = \frac{1}{N}\sum_i (x_i - \bar{x})^2 + (\frac{1}{N}\sum_j x_j - X)^2$$

$$= \cdots$$

$$= \frac{1}{N}\sum_i (x_i - \bar{x})^2 + \frac{\sigma^2}{N} + \frac{\sum_{j\neq k}(x_j - X)(x_k - X)}{N^2}$$

標準偏差の最良推定値

σ²を解けば

$$\sigma^2 = \frac{1}{N-1}\sum_i (x_i - \overline{x})^2 + \boxed{\frac{\sum_{j \neq k} (x_j - X)(x_k - X)}{N(N-1)}}$$

N個の計測データをあつめ、
σ²の計算を何回も実行すると、
この項は平均的に0となる

何故か？ ⇒ 計測値xは真の値Xのまわりに正規分布し、各計測値が独立ならば、上記の積は正負がひとしく出現して打ち消しが起こるから

不偏分散

同一のものに対する複数の計測値が正規分布に従う場合、計測誤差の分散の最良推定値は

$$\sigma^2 = \frac{1}{N-1}\sum_i (x_i - \overline{x})^2$$

で与えられる。これを**不偏分散**という。

⇒ 不偏分散の平方根は
計測一回あたりの誤差の見積もりになっている

不偏分散の意味

$$\sigma^2 = \frac{1}{N-1}\sum_i (x_i - \overline{x})^2$$

不偏とは？ ⇒偏らないということ。
分母がNだと計測誤差の真の分散より小さいほうにずれるため、そのずれがない推定量だということ

偏らない⇒ 統計的には何度も試行したときに平均が真の値と一致すること。
⇒ **不偏推定量**という。

計測誤差が正規分布する場合、複数の計測値からもとめた最尤推定分散は、不偏推定量であることがわかる

二乗和を使う根拠

- ランダムで独立な誤差の場合、誤差の伝播は二乗和を用いた。
  しかし、その理由はまだ述べていなかった。
- 正規分布する誤差の場合にこれが説明できることをこれから述べる

測定値と固定値の和の分布

- q=x+A という量を考える

  x：平均X、標準偏差σxの正規分布に従う

  A:固定値

$$p(x) \propto \exp\left(-\frac{(x-X)^2}{2\sigma_x^2}\right) \Rightarrow p(q) \propto \exp\left(-\frac{(q-A-X)^2}{2\sigma_x^2}\right)$$

qはX+Aを中心に標準偏差σxの正規分布に従う
⇒固定値との和は平均がシフトするだけ
誤差は変化なし

測定値と固定値の積の分布

- q=Bxの分布

  x：平均X、標準偏差σxの正規分布に従う

  B:固定値

$$p(x) \propto \exp\left(-\frac{(x-X)^2}{2\sigma_x^2}\right) \Rightarrow p(q) \propto \exp\left(-\frac{(q-BX)^2}{2B^2\sigma_x^2}\right)$$

qはBXを中心に標準偏差Bσxの正規分布に従う
⇒固定値との積は平均と誤差がともに
固定値倍される
⇒c.f.誤差伝播式を思い出すこと

二つの測定値の和の分布

- x： 平均X,標準偏差σxの正規分布
- y： 平均Y,標準偏差σyの正規分布

  x,yはともに独立とする。

q＝x＋yという量を考えると、

$$p(x) \propto \exp\left(-\frac{(x-X)^2}{2\sigma_x^2}\right)$$

$$p(y) \propto \exp\left(-\frac{(y-Y)^2}{2\sigma_y^2}\right)$$

$$p(x,y) \propto \exp\left(-\frac{(x-X)^2}{2\sigma_x^2}\right)\exp\left(-\frac{(y-Y)^2}{2\sigma_y^2}\right)$$

二つの測定値の和の分布

$$p(x,y) \propto \exp\left(-\frac{(x-X)^2}{2\sigma_x^2}\right)\exp\left(-\frac{(y-Y)^2}{2\sigma_y^2}\right)$$

$$= \exp\left(-\frac{(x-X)^2}{2\sigma_x^2} - \frac{(y-Y)^2}{2\sigma_y^2}\right) \qquad z^2 = \frac{\left[\sigma_y^2(x-X)-\sigma_x^2(y-Y)\right]^2}{\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)}$$

$$= \exp\left(-\frac{(x+y-X-Y)^2}{2(\sigma_x^2+\sigma_y^2)} - \frac{z^2}{2}\right)$$

$$= \exp\left(-\frac{(x+y-X-Y)^2}{2(\sigma_x^2+\sigma_y^2)}\right)\exp\left(-\frac{z^2}{2}\right)$$

zで積分すれば、
x＋yの確率がでる

二つの測定値の和の分布

$$p(x+y) \propto \exp\left(-\frac{\{x+y-(X+Y)\}^2}{2(\sigma_x^2+\sigma_y^2)}\right)$$

結局q=x+yは、平均X+Y, 分散σx²+σy²の正規分布に従うことがわかる

誤差（標準偏差）は $\sqrt{\sigma_x^2+\sigma_y^2}$

たしかに二乗和で誤差が伝播していることがわかる

一般の関数の場合の分布

q(x,y)の分布を求める
x、yがそれぞれ平均X,Yの近くにあるとき以下の近似が成立（テーラー展開）

$$q(x,y) \approx q(X,Y) + \left.\frac{\partial q}{\partial x}\right|_{x=X}(x-X) + \left.\frac{\partial q}{\partial y}\right|_{y=Y}(y-Y)$$

$q(X,Y)$ ： 固定値なので平均がシフト

$\left.\frac{\partial q}{\partial x}\right|_{x=X}(x-X)$ 　(x－X)は分散σx²で正規分布
さらに偏微分は固定値倍で効く

一般の関数の場合の分布

$$q(x,y) \approx q(X,Y) + \left.\frac{\partial q}{\partial x}\right|_{x=X}(x-X) + \left.\frac{\partial q}{\partial y}\right|_{y=Y}(y-Y)$$

qは平均が $q(X,Y)$

標準偏差 $\sqrt{\left(\left.\frac{\partial q}{\partial x}\right|_{x=X}\right)^2 \sigma_x^2 + \left(\left.\frac{\partial q}{\partial y}\right|_{y=Y}\right)^2 \sigma_x^2}$

の正規分布に従う

＊多変数関数における誤差の式と一致する

標本平均値の分布

- 標本平均 $\bar{x} = \frac{x_1+\cdots+x_N}{N} = \frac{1}{N}\sum_i x_i$
- $x_1,\ldots x_N$はそれぞれ真の値Xのまわりに標準偏差σxで正規分布している
  すると、標本平均の平均と誤差（標準偏差）誤差伝播の式を使って見積もることができる

  標本平均はx_iの和と定数倍から計算されるので、やはり正規分布することがわかる

標本平均値の分布

- 標本平均 $\bar{x} = \frac{x_1+\cdots+x_N}{N} = \frac{1}{N}\sum_i x_i$
- 標本平均の平均 $\bar{\bar{x}} = \frac{X+\cdots+X}{N} = X$
- 標本平均の分散

$$\sigma_{\bar{x}} = \sqrt{\left(\frac{\partial \bar{x}}{\partial x_1}\sigma_{x_1}\right)^2+\cdots+\left(\frac{\partial \bar{x}}{\partial x_N}\sigma_{x_N}\right)^2} = \sqrt{\left(\frac{\sigma_{x_1}}{N}\right)^2+\cdots+\left(\frac{\sigma_{x_N}}{N}\right)^2}$$

$$= \sqrt{N\left(\frac{\sigma_x}{N}\right)^2} = \boxed{\frac{\sigma_x}{\sqrt{N}}}$$

← 標本平均の分散は計測値の個数Nの平方根に反比例

標本平均値の分布

標本平均 $\bar{x} = \frac{x_1+\cdots+x_N}{N} = \frac{1}{N}\sum_i x_i$

は平均が真の値 X

標準偏差が $\frac{\sigma_x}{\sqrt{N}}$ の正規分布に従う

xが正規分布するときはもちろん、xが別の分布に従うときでも
Nが十分大きいときの標本平均は正規分布する

中心極限定理：同一分布に従う多数の和は正規分布する

標本から真の値の分布を推定

- 標本 {x1,…xN} から真の値Xの分布を推定する

最良推定量 $\bar{x} = \frac{x_1+\cdots+x_N}{N} = \frac{1}{N}\sum_i x_i$

真の計測誤差 σ はわからないので不偏分散で推定

標準偏差 $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} = \sqrt{\frac{\sum_i (x_i-\bar{x})^2}{N(N-1)}}$

真の値Xは $\bar{x} \pm \sigma_{\bar{x}}$ の中に68%の確率で存在するといえる

ここでチェック：

- ある量xを6回測定して以下の結果をえた
  51,53,54,55,52,53

(a) 標本の平均値と標本の標準偏差：
平均値＝53, 標準偏差＝√(10/6) = 1.29

(b) xの測定における1回あたりの誤差の見積もり
不偏分散の平方根で評価する
√(10/5)=√2＝1.41

(c) 真の値Xの最良推定値と標準偏差
最良推定値は標本平均：53
標準偏差は平均値の標準偏差を求める
⇒不偏分散の平方根を√Nで割ったもの
√2/√6 ＝ √(1/3) = 0.58