物理実験学

竹下徹

数値データと真の値の関係

我々の測定は、ある量の正しい本当の値＝真の値があり、これを見つけることを目的に行われる。　　しかし、測定の回数は有限であり、そのたびごとに 異なる（もちろん測定装置の精度内で一致することもあるが）のが常識である。
では真の値はこの有限回で異なる数値の測定からどうやって導くべきか？

ある種の信念かも知れないが、ここでは、法則があると思う。それは
有限回の測定をつづけ無限回に達したとき得られる答えが、真の値である。
すなわち、　有限回(N回)の測定から得られた平均値、や　標準偏差　は
　N->∞　で真の値とその分布の標準偏差と一致する。

これを「大数の法則」という。

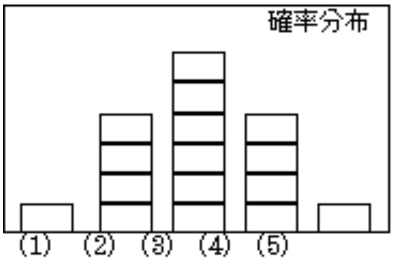
真の値が測定の結果となるべきである、は当然である。
しかし、分布とはなんだ！？　　真の値を得る過程で測定の複数回のくり返しは、有限の標準偏差＝ひろがりを持つ。-----ただしどういう形の分布である かは、今は分からない。

で、分布とは、
例として、　　４枚のコイントスを考える。ただし、コインの区別はつかないとしよう。とすると、結果は４種類に分類される。この場合、離散型のデータなの で状態（結果）は有限の種類となる。その5種は、４枚のうち、
(1)４枚とも、H, (2) 3枚H,(3)2枚H,(4)1枚H,(5)0枚H＝この場合全部Tということ。
ただし、分布は、この種類のそれぞれの棒グラフの高さ（回数）をグラフにして表される。ーーーー＞実験せよ！そして有限回の分布を示せ。

仮定：すべてのコインはHとTのでる確率はそれぞれ1/2=0.5
(1)４枚とも、H　：場合は、１通り、(HHHH) この確率＝P(4)=(1/2)4=1/16=0.0625
(2) 3枚H　：　　場合が４通り＝(HHHT,HHTH,HTHH,THHH) P(3)=4*(1/2)4=4/16=0.25
(3)2枚H：　場合は、６通り=(HHTT,HTHT,HTTH,THHT,THTH,TTHH)
P(2)=6*(1/2)4=6/16=0.375
(4)1枚H：　場合が４通り＝(TTTH,TTHT,THTT,HTTT) P(1)=4*(1/2)4=4/16=0.25

(5) 1枚H １通り、(TTTT) この確率＝P(0)=(1/2)4=1/16=0.0625

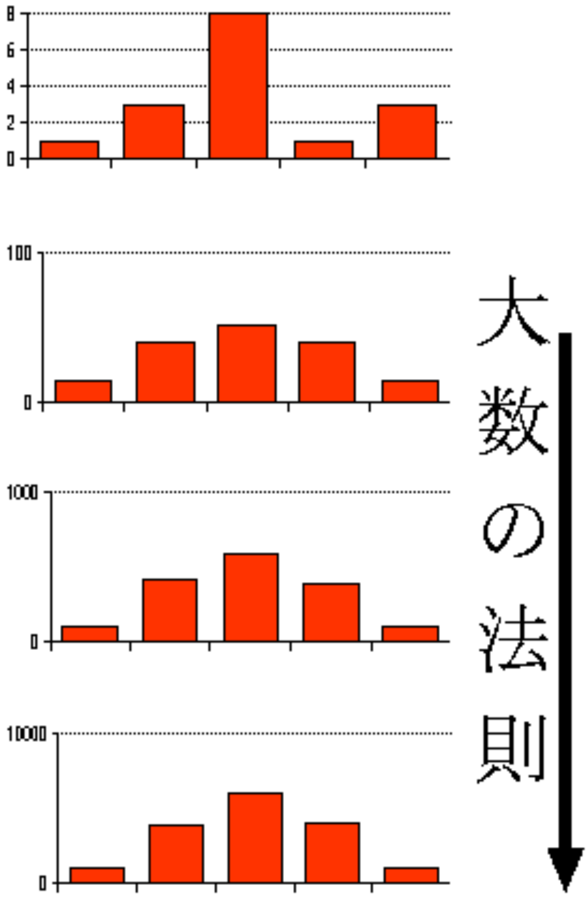
全確率は１でなくてはならない。P(0)+P(1)+P(2)+P(3)+P(4)=
0.0625+0.25+0.375+0.25+0.0625=1.000



確率分布と確率変数（整数型）

分布とは、 離散的な場合
例として、　　４枚のコイントスを考える。ただし、コインの区別はつかないとしよう。とすると、結果は４種類に分類される。この場合、離散型のデータなの で状態（結果）は有限の種類となる。この種類を表す下図を確率変数という。確率変数　5種は、４枚のコインのうち、(1)４枚とも、H, (2) 3枚がH,(3)　2枚がH,(4)　1枚がH,(5)　0枚がH =この場合全部Tということ。ただし、分布は、この種類のそれぞれの棒グラフの高さ（回数）をグラフにして表される。
実験してみる。16回ー４つのコイントスを行った結果、Tが出た回数(0)1回,(1)3回,(2)8回,(3)1回,(4)3回、160回ー４つのコイ ントスを行った結果、
Tが出た回数(0)14回,(1)40回,(2)51回,(3)41回,(4)14回、1600回ー４つのコイントスを行った結果、Tが出た回数(0) 97回,(1)414回,(2)586回,(3)397回,(4)106回、16000回ー４つのコイントスを行った結果、Tが出た回数 (0)988回,(1)3935回,(2)5927回,(3)4094回,(4)1056回。

回数を増やすほど理想的な分布に近づくことが想像できる。無限回の試行は理想分布に一致するはずというのが、大数の法則である。横軸に確率変数を取 り、縦軸に確率を取ったグラフがつぎである。



0	T	H	H	T	2
1	T	H	T	H	2
2	H	H	H	H	0
3	H	H	T	T	2
4	T	T	T	T	4
5	T	H	T	H	2
6	H	T	T	H	2
7	T	T	T	T	4
8	T	T	T	T	4
9	H	H	H	T	1
10	H	H	T	H	1
11	H	T	T	H	2
12	T	T	T	H	3
13	H	T	H	T	2
14	H	T	H	F	1
15	H	T	T	H	2
16	1	3	8	1	3

期待値とは、

例として、　　４枚のコイントスを考える。可能な状態は次の５通りである。すなわち４枚のうち、
(1) ４枚とも、H, (2) 3枚がH,(3)　2枚がH,　(4)1枚がH,　(5)0枚がH =この場合全部Tということ。
各の確率は、P(1)=0.0625(1/16),P(2)=0.25(4/16),P(3)=0.375(6/16),P(4)=0.25 (4/16),
P(5)=0.0625(1/16)であることは計算した。
ここで、１回だけ試したとき出る可能性r枚表の期待値<r>を次のように定義する。いままでxiというデータの平均を取る作業と同じことを、確率という重みを付けてとることに相当する。

$$< r >= \sum_r rP(r) = 0P(0) + 1P(1) + 2P(2) + 3P(3) + 4P(4)$$

$$= 0 + \frac{4}{16} + 2\frac{6}{16} + 3\frac{4}{16} + 4\frac{1}{16} = \frac{4 + 12 + 12 + 4}{16}$$

$$= \frac{32}{16} = 2$$

この<r>は表の回数の期待値 と呼ばれる。 これは必ずしも確率の最大の値ではない。（ここでは、偶然一致しているが、、一般には　<r>は整数とは限らない。データの代表値と分散あるいは標準偏差を議論したように

データの広がりを表す。

$$S^2 = \langle r^2 \rangle - \langle r \rangle^2$$

$$= \sum r^2 P(r) - \left(\sum r P(r)\right)^2$$

<r2>は２乗平均であり

$$\langle r^2 \rangle = \sum r^2 P(r) = 0 + 1^2 \frac{4}{16} + 2^2 \frac{6}{16} + 3^2 \frac{4}{16} + 4^2 \frac{1}{16}$$

$$= \frac{0+4+24+36+16}{16} = \frac{80}{16} = 5$$

よって

$\sigma^2 = \langle r^2 \rangle - \langle r \rangle^2$ =5-4=1　となる。

確率分布：連続変数の場合の連続分布

次に　rが連続あるいは実数値を取る場合を考える。同じ計算ではまず事は明らかである。
例えば棒の長さを測ったとき、（長さxは連続値を取りうる）
長さが11cmという意味を考えよう。これは、最小メモリが10cmの物差しで測ったなら、
10cm〜12cmの意味（ここで〜は、その間ｬのどこかという意味）である。
　しかし、最小メモリが1cmの物差しで測ったなら、　10.5cm〜11.5cmの意味
　　　　　最小メモリが1mmの物差しで測ったなら、　10.95cm〜11.05cmの意味
のように色々な言い方が可能である。（有効数字をさておいて）

そこでも確率を定義する、Aただし上の例のように連続量は、ある値からある値の間にあることを持って、話しをする。すなわち　測定量xがx1からx2の間にある確率を
P(x1,x2)として与える事のみができる。ただし、次の式を考え $\hat{P}(x)$ を本質と考える。

$\hat{P}(x)$ を確率密度関数と呼ぶ。 $P(x1, x2) = \int_{x1}^{x2} \hat{P}(x)dx$ $\hat{P}(x) = \lim_{\delta x \to 0} \frac{P(x, x+\delta x)}{\delta x}$ こう言っても良い。
注意 確率P(x)は無次元量であるので、

$\hat{P}(x)$ はｘの次元の逆数である：$[\hat{P}(x)] = \frac{1}{[x]}$
さてこれでようやく、xの期待値を計算する準備ができた。

　xの期待値<x>は離散的な場合と同様にｘのその期待値をかけて和を取ればよい、が

xは連続量なので、積分を用いる。 $< x >= \int_{-\infty}^{\infty} x\hat{P}(x)dx$ 同様に標準偏差 σ^2 は

$$\sigma^2 =< x^2 > - < x >^2$$

$$= \int x^2 \hat{P}(x)dx - (\int x\hat{P}(x)dx)^2$$ と かける。 $\hat{P}(x)$ の具体例は次の章で示す。

宿題：離散的確率分布する場合の例を２つ以上考え、その期待値と標準偏差を計算せよ。
　　　実際可能なら実験し、その理論分布と実際を比較せよ。
　　　ヒント：一つは、一様分布と呼ばれ、サイコロや転がり鉛筆の場合である。

物理実験学　　24/April/2003 質問表まとめ

　正しい字：乱数

質問：「<r2>-<r>2で広がりを求めていますがどういう意味があるのですか？」
答え：　データxiが与えられてその平均値x。と二乗平均　$\overline{x^2}$ が計 算されたとき標準偏差 $\sigma^2 = \frac{1}{N}\sum(x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$ が広がりを表すように、確率が与えられた場合 <r2>-< r>2が確率分布の広がりを表す量となる。

$$S^2 =< r^2 > - < r >^2 = \sum r^2 P(r) - \left(\sum rP(r)\right)^2$$ と

$$\sigma^2 = x^2 - \overline{x}^2 = \sum x^2 \frac{1}{N} - \left(\sum x \frac{1}{N}\right)^2$$ はP(r)=1/N

で同じにみえませんか？
大数の法則を使って多数の測定から「真の価」を知りたいとき測定は確率で表現されるため、この形を使います。ヒストグラムを使って積みあげてゆくと全体数 で各ビンの高さを割った値つまり確率を用いることになります。従って両者は異なる物ではなく、同一と考えられましょう。

　　また重み付き平均を使うと、 $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$ 、ここでwiの代わりに確率P(xi) を使い、確率の和は１となるべしという要請をおけば、分母は自動的に１となり式は $\bar{x}_w = \sum x_i P(x_i)$ と なります。これはxiを ｒ に置きかえれば同じ式ですす。このように同じ式も変数名を変えると違う式に見ないように練習してください。

物理実験学　　24/April/2003 質問表まとめ

正しい字：乱数

質問：「$\sigma^2 = \frac{1}{N}\sum(x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$ で広がりを求めていますがどういう意味があるのですか？」

答え：　データxiが与えられてその平均値x。と二乗平均 $\overline{x^2}$ が計 算されたとき標準偏差 $\bar{x}$ が広がりを表すように、確率が与えられた場合 <r2>-< r>2が確率分布の広がりを表す量となる。

$S^2 = \langle r^2 \rangle - \langle r \rangle^2 = \sum r^2 P(r) - \left(\sum rP(r)\right)^2$ と

$\sigma^2 = x^2 - \bar{x}^2 = \sum x^2 \frac{1}{N} - \left(\sum x \frac{1}{N}\right)^2$ はP(r)=1/N

で同じにみえませんか？

大数の法則を使って多数の測定から「真の価」を知りたいとき測定は確率で表現されるため、この形を使います。ヒストグラムを使って積みあげてゆくと全体数 で各ビンの高さを割った値つまり確率を用いることになります。従って両者は異なる物ではなく、同一と考えられましょう。

　また重み付き平均を使うと、$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$ 、ここでwiの代わりに確率P(xi) を使い、確率の和は１となるべしという要請をおけば、分母は自動的に１となり式は $\bar{x}_w = \sum x_i P(x_i)$ と なります。これはxiを r に置きかえれば同じ式です。このように同じ式も変数名を変えると違う式に見ないように練習してください。

物理実験学　　30/April/2003 質問表まとめ

質問：「確率密度関数 $\hat{P}(x)$ の意 味がわかりません？またなぜ妙な単位を持つのですか？」

答え：　連続量に対する確率を定義する事ができません。出現する場合の数が有限な離散分布（例えばサイコロの目の出現確率分布は６個の場合の数しかありませ ん）の場合はそれぞれの場合を現す数 r（サイコロの目ならr=1,2,3,4,5,6)について対応する確率 $P(r)$ を定 義できる。しかし連続量では出現する場合の数は無 限大　（∞　infinite）です、その場合出現する場合を現す変数　x　の関数である確率をどうすれば定義できるでしょうか？そこで確率 密度は $\hat{P}(x)$ を導入して、連続量xが x1< x < x2　にある確率は原理的に定義可能で、その確率を　　次のよ

うに定義しよう。 $P(x.x1 < x < x2) = \int_{x1}^{x2} \hat{P}(x)\,dx$ 従って自動的に確率密度関 数 $\hat{P}(x)$ はdxをかけて確率になる し、確率は単位がないので、$\hat{P}(x)$ は dxの単位の逆数となります。この定義はいままで離散的な場合の数の場合に定義し他方法と全く類似して定義できます。
全て左の連続量と右の離散量が対応していることに注意してください。

質問2：「重み付き平均の例はありますか？」

　重み付き平均を使うと、 $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$ 、 と書かれます。サイコロの目(1,2,3,4,5,6)をxiが取るとき、複数回サイコロを振って（その目の値をyiとしよう）その平均は $\bar{y} = \sum_{i=1}^{N} y_i$ とでますが、これをヒストグラム化して、yiは必ず１ から６までの整数ですからxiにヒストグラムとします。wiは目xiが出た回数となり、平均は重み付き平均となります。つまりヒストグラムの場合エント リーしたそれぞれを１と数えるならwiは１の何倍のエントリーがあったのかと言うことになり、全体の平均はwiの重みを付けて重み付き平均を求めることに より計算できます。

質問3：「有効数字で四捨五入するときはありますか？」
有効数字とは測定できて信頼できる数字、または精度があると考えられる数字のことです。たとえば　1mmまで測定できる物差しで測った長さを最小目盛りの 1/10まで読めというのは、有効数字が小数点下１桁（mm単位で）であり、最後桁の値は＋１あるいはー１ずれるかもしれないとおィもって使うことを意識 してください。正し計算の途中では四捨五入を繰り返さない方が結果に信頼性が生まれます、なぜならに途中で減算（引き算）が含まれる 場合は精度がわるくなる可能性が高いからです。最終結果は必ず有効数字で表してください。つまり途中は１桁大きく計算してきた結果を四捨五入して有効数字 に丸めてください。しかし、標準偏差を計算してみると、ここでは減算 $<x^2> - <x>^2$ が 効いてくるので、ほとんど ゼロに近くなり、有効数字が現れないことがあります、例えば $<x> = 10.23, \sigma = 0.0024$ こういう時は標準偏差を小 数点下２桁にするべきですが、丸めると $\sigma = 0.00$ とな ります。Bこの値は精度を表す量なので切り上げ標準的です、従って $\sigma = 0.01$ と答 えてください。

質問4：「πの有効数字は何桁ですか？」
πの値や，物理定数を使う場合には，計算する最大桁の有効数字よりも1桁以上大きな値を使って計算してください。例えば

$4\pi r2 = 4\times3.1416\times(2.674m)2 = 8.985m2$　　ここでπは5桁とりました、なぜなら2.674に４桁あるからです。また上の計算で，最初の４は1桁の有効数字というわけではありません、4.0000と扱っています。
