# 誤差と不偏分散

## 1　誤差

### 1.1　誤差の定義

測定において，得られた測定値には必ず何らかの不確かさが伴う．そして，真の値を知ることはできない．このような測定の操作の中で生じる不確かさは**誤差**と呼ばれる．誤差は測定値から真の値を引いた値である．誤差の真の値に対する比を誤差率または相対誤差という．しかしながら，真の値というものは観念的な量であり，計測器や測定方法をどんなに改良し，注意深く測定しても実際には求められない．それでも，できるだけ正確に測定を行い，また，多くの測定値の平均値を取ることによって，真の値に近づくことは可能である．すなわち，繰り返し測定することで，誤差の近似値を**推定する**ことができる．

測定値を統計量として扱う立場では，測定値（誤差）はすべて確率的なものであって，ばらつき（分布）を持っており，その統計的な性質を明らかにすることを測定の目的とする．測定とは図 1 に示すように，無限にある測定値の候補の中からランダムにどれか一つを選ぶ作業ということになる．そのような測定値の候補の集合を**母集団**といい，抽出した測定値を**標本**と呼ぶ．

### 1.2　誤差の種類

誤差は大きく分けて，系統誤差，偶然誤差，そして，間違いによる誤差の 3 つに分類される．

(1) 間違いによる誤差

測定者の不注意や間違い，記録の誤り，測定手続き上の過失などにより測定値に含まれる誤差である．

(2) 系統誤差

系統誤差とは測定結果のかたよりの原因となる誤差のことであり，測定器の目盛りの誤差のように常に一定なかたよりを与えるものや，測定の方法に適用した理論や仮定の不完全さにより生じるもの，環境条件の変化により計測器や測定量が規則的な変化を起こすことによる誤差，目盛りの読み取りや計測の調整などの個人誤差，などがあげられる．この種の誤差は，原因がわかれば誤差の大きさを計算して補正することが可能である．しかし，一般的に系統誤差の大きさはすべての測定値に一定のかたよりを与えるので，測定結果からは推定できない．測定条件や理論から推定することや，測定条件や測定装置，もしくは，測定方法を変えて測定を試み，誤差の原因を知る手がかりを探すなどするしかない．

(3) 偶然誤差

間違いによる誤差や系統誤差をすべて取り除いても，なお多数の原因が互いに独立かつ不規則に作用して生じる誤差がある．これを偶然誤差（確率誤差）という．この誤差の原因はつきとめられないので，どうしても測定値にはばらつきが生じる．大きさも符号もランダムに生じるので，この誤差を扱うには統計的な処理が必要となる．これまでの経験から偶然誤差のばらつきは**正規分布**に従うことが知られている．そこで，測

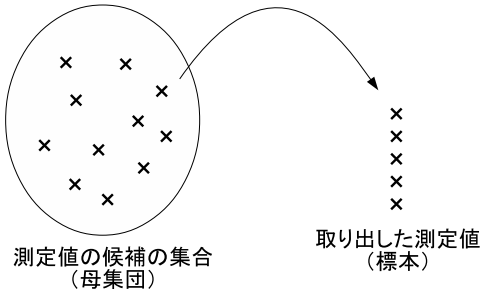


図 1: 測定における母集団と標本

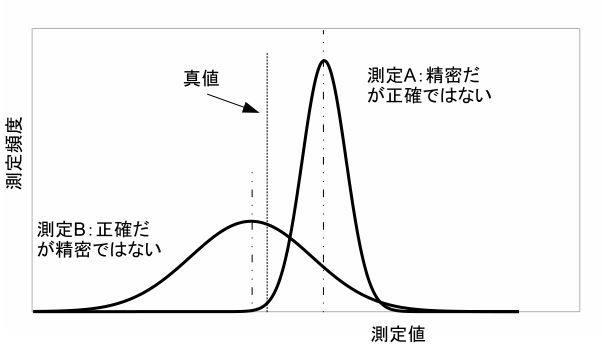


図 2: 正規分布に従う測定値のばらつき

定の結果は図 2 に示すようなものになることが予想される．ここで，横軸は測定値であり，縦軸はその測定値を得られる頻度（確率密度）に対応している．測定値の分布の平均（母平均）$m$ は真の値 $T$ とは一致しないので，その違い $\delta$ をかたより（系統誤差）と考える．そして，ある測定値 $x$ と母平均 $m$ との差を偏差 $\varepsilon$ という．母平均 $m$ を知るためには無限回数の測定を行う必要があるが，通常それは実現できないので，その近似値として標本平均 $\bar{x}$ を使用する．また偏差も求めることができないために，ばらつきの度合いを知るために $x - \bar{x}$ の値を用いるが，これを**残差** と呼ぶ．

### 1.3 精度

測定においてはしばしば精度という言葉がよく聞かれる．実は，精度については測定上の定義は無い．計測の分野では**正確さ**と**精密さ**という二つの指標で測定の精度を評価する．かたより（系統誤差）の少ない程度を正確さといい，ばらつき（偶然誤差）の小さい程度を精密さという．図 2 に示すばらつきのかたちをみると，測定 A は真の値から標本平均は外れているがばらつきの度合いが小さいので精密な測定であり，測定 B はばらつきは大きいものの正確な測定を示す．

## 2 不偏分散

### 2.1 ばらつきの数値化

前述のように，測定値の母平均 $m$ は求めることができないので，測定の実験では残差がばらつきを表す元の量となるが，標本平均の周りにばらついている残差を加え合わせても正と負の値を持つために相殺して確かな値とはならない．そこで，以下のように考える．

$n$ 個の測定値（$x_1,\ x_2,\ \cdots,\ x_n$）の平均値 $\bar{x}$ $(\bar{x} = \sum x_i/n)$ を用いて，

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

のように残差の 2 乗和 $S$ を求める．この $S$ のことを平方和と呼ぶこともある．

### 2.2 標本分散

平方和の大きさは測定値の個数に依存するので，平方和を $n$ で割ることで個数の影響を排除できる．そこで，

$$s^2 = \frac{S}{n} = \frac{\sum (x_i - \bar{x})^2}{n} \tag{2}$$

で求められる $s^2$ を標本分散といい，ばらつきの目安として使用できる．またその平方根 $s$ を標本標準偏差という．

## 2.3 不偏分散

測定を行う際に知りたいのは真の値であるが，それは決して知ることができないために，次に知りたい候補としては母平均 $m$ である．しかし，それも知ることは困難であるので，推定値がその周辺にどの程度ばらついているのかを知るために分散を求めている．しかし，前述の標本分散は母分散とは一致しないことが知られている．これは残差を求める際に標本平均値を用いているために，$n$ 個の測定値の自由度が $n-1$ となってしまっていることによる．そこで，平方和を $n-1$ で除したものを不偏分散と呼び，使用することが一般的に行われている．この不偏分散は母分散の良い推定値であることがわかっている．

$$v^2 = \frac{S}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1} \tag{3}$$

ここで，本授業では不偏分散の記号として $v^2$ を用いることとするが，統計学の分野では，実は分散に関する記号は統一されておらず，本によりまちまちで表現されている．アルファベットの $v$ が使われている理由は分散のことを英語で variance と呼ぶためである．なお，表計算ソフトのエクセルなどで分散を求める関数は `var` となっているが，これは不偏分散を求める関数である．同様に，標準偏差を求める関数 `stdev` は不偏分散の平方根 $v$ であり，standard deviation の略である．不偏分散の平方根に対する呼び名も実は決まっていない．ここでは便宜上「標準偏差」と呼ぶときは不偏分散の平方根であることに注意すること．表 1 に呼び名と記号を整理しておく．

表 1: 平均，分散と標準偏差の記号

| | | | |
|---|---|---|---|
| 母平均 | $m$ | 標本分散 | $s^2$ |
| 母分散 | $\sigma^2$ | 標本標準偏差 | $s$ |
| 母標準偏差 | $\sigma$ | 不偏分散 | $v^2$ |
| 標本平均 | $\bar{x}$ | 不偏分散の平方根（標準偏差） | $v$ |

## 2.4 平均値の分散

標本平均値は母平均を推定する上でよい近似値として知られているが，標本の選び方は無数にあるといえるので，標本平均値も複数個求めることができる．すなわち，繰り返し測定を行って一つの平均値を出すことも可能であるが，何度かの測定をまとめていくつかの平均値を求めることもできるということである．このとき，$n$ 個の平均値の不偏分散は 1 つの平均値から出した不偏分散の $1/n$ となることが知られている．

$$v^2(\bar{x}) = \frac{v^2}{n} \tag{4}$$

また，元の測定値が正規分布に従わない場合でも平均値 $\bar{x}$ の分布は正規分布となることが知られている．

## 2.5 平方和の計算式

式 (1) で平方和は求めることができるが，これは実際に計算する際には手続きが面倒である．そこで，電卓などでも計算しやすくするための変形が昔から知られている．以下の式と $S$ が等価であることを確認しておくこと．

$$\begin{aligned} S &= \sum x_i^2 - n\bar{x}^2 && (5)\\ &= \sum x_i^2 - \bar{x}\sum x_i && (6)\\ &= \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 && (7) \end{aligned}$$

## 3　二乗平均誤差

標準偏差と同じ定義の以下の式，

$$\varepsilon_m = \sqrt{\frac{\sum (x_i - m)^2}{n}} \tag{8}$$

を二乗平均誤差という．標準誤差と呼ばれることもある．英語表現の Root mean square から RMS 誤差と呼ばれることもある．ここでも，標本平均を用いるのが通常であるので，不偏分散の平方根を取り，

$$\varepsilon'_m = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \tag{9}$$

の式の方が推定値として用いられることが多い．