自己相関関数

1　平均値

■40 と 60 の平均値は 50 である。■10 人のうち 5 人の体重が 40kg で，残りの 5 人が 60kg なら平均値は 50kg である。■もし，10 人のうち，3 人の体重が 50kg で残りの 7 人が 60kg なら，全員の体重の合計を全員の人数で割って

$$\text{平均体重} = \frac{3 \times 50 + 7 \times 60}{10} = \frac{3}{10} \times 50 + \frac{7}{10} \times 60 = 57\text{kg}$$

である。■車が $T = 10$ 時間のうち，前半の 5 時間は 40 km/h で走り，後半の 5 時間は 60 km/h で走った。 平均速度は $\overline{v}$ =50km/h である。■もし，車が $T = 10$ 時間のうち，初めの $\Delta t_1 = 3$ 時間は $v_1 = 40$km/h で走り，残りの $\Delta t_2 = 7$ 時間は $v_2 = 60$km/h で走った。 この場合には時間の長さが違うので，時間の割合を考えて平均値は次のようになる。

$$\overline{v} = \frac{3 \times 40 + 7 \times 60}{10} = 57 \text{ km/h}$$

以上のような平均値の計算は日常生活でいつも使っている。 最後の例を文字式で表すと

$$\overline{v} = \frac{\Delta t_1 \cdot v_1 + \Delta t_2 \cdot v_2}{T}$$

である。 Δt_1，Δt_1 は次に述べる重み p_1，p_2 に相当する。

1.1　重み付きの平均値計算

上の例では，$v_1 = 40$km/h は全体の時間 10 のうち 3 の割合で平均速度に寄与し，$v_2 = 60$km/h は 10 のうち 7 の割合で平均速度に寄与している。平均値はこの 2 つの寄与の和である。 このとき 3 と 7 を**重み**と呼ぶ。

一般的な重み付きの平均値計算は次のようなものである。 v を測定したとき，値 v_1 が p_1 回，v_2 が p_2 回，v_3 が p_3 回，$\cdots$ 得られたとき，p_i を重み（weight）と呼び，平均値は次式で与えられる。

$$\overline{v} = \frac{p_1 v_1 + p_2 v_2 + p_3 v_3 + \ \cdots}{p_1 + p_2 + p_3 + \cdots} = \frac{\sum p_i v_i}{\sum p_i} \tag{1}$$

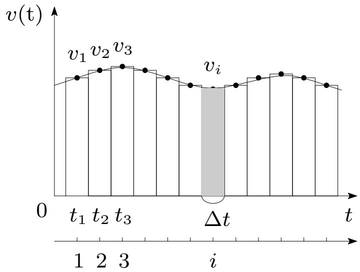
1.2　統計データの平均

以下ではもう少し複雑な場合の平均値を考える。 時間的に変化する電圧を連続して測定したとしよう。 データ解析の基本は平均値である。この場合には，多数のデータについての統計的な平均値を求める計算が必要である。

一定の時間間隔 $\varDelta t$ で速度が $v_1, v_2,\ \cdots\ , v_N$ と変化した場合の平均速度は式（1）で

$$\overline{v} = \frac{v_1\varDelta t + v_2\varDelta t + \ \cdots\ + v_N\varDelta t}{\varDelta t + \varDelta t + \ \cdots\ + \varDelta t} = \frac{\displaystyle\sum_{i=1}^{N} v_i\varDelta t}{\displaystyle\sum_{i=1}^{N}\varDelta t} = \frac{\displaystyle\sum_{i=1}^{N} v_i\varDelta t}{N\cdot\varDelta t} \tag{2}$$

$$\overline{v} = \frac{1}{T}\sum_{i=1}^{N} v_i\varDelta t \qquad T = N\varDelta t \text{ は全体の時間} \tag{3}$$



である。 式（2）の分子の $\varDelta t$ は式（1）の重み p_i であり，全て同じ値 $\varDelta t$ である。 i 番目の速度を v_i として Σ の記号を使えば式（2）第 3 項の形に表すことができる。 この式の $v_i\varDelta t$ は時間を横軸，速度を縦軸にとって表した棒グラフの i 番目の棒の面積で，$\varDelta t$ の間に進んだ距離である。 全ての棒の面積の合計は全時間 T の間に進んだ総距離であり，T で割れば平均速度が得られるというあたりまえのことになる。

1.3　連続量の時間平均

時間の区間を短くし，$\varDelta t \to 0$ の極限をとれば，v_i は t の連続的な関数 $v(t)$ になり，Σ は積分に置き換えられて，時刻 0 から T の間の平均速度は

$$\overline{v} = \frac{1}{T}\sum_{i=1}^{N} v_i\varDelta t \quad \to \quad \overline{v} = \frac{1}{T}\int_0^T v(t)\mathrm{d}t$$

と表される。 車が永遠に走り続ける場合，平均速度は

$$\overline{v} = \lim_{T\to\infty}\frac{1}{T}\int_0^T v(t)\mathrm{d}t \tag{4}$$

と表すことができる。 これは連続関数の平均値の一般的な形である。

1.4　アンサンブル（集合）平均

全く同じ測定器が M 台あるとする。 M は大きな数である。装置 1，装置 2，$\cdots$ ，装置 M から共通な時間 t の関数として $x_1(t)$, $x_2(t)$，$\cdots$ ，$x_M(t)$ のデータを得たとする。時刻 t の瞬間に同時に得られる全装置の測定値の平均は，s 番目 の装置の信号を x_s として，

$$\hat{x}(t) = \frac{1}{M}\sum_{s=1}^{M} x_s(t)$$

である。 装置の台数を限りなく多くした場合は

$$\hat{x}(t) = \lim_{M\to\infty} \frac{1}{M}\sum_{s=1}^{M} x_s(t)$$

である。 このような平均をアンサンブル（集合）平均と呼ぶ。

1.5　エルゴード的

どの瞬間においても，**アンサンブル平均が変化せず一定である信号は，定常的である。** ランダムな（無秩序な）信号は時間的に性質が変化せず定常的である。 定常的な信号の場合には，ある 1 台の装置で長時間測定した信号の時間平均

$$\overline{x} = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\mathrm{d}t$$

と，多数の台数の測定器で，ある瞬間 t に一斉に測定した信号の**アンサンブル（集合）平均**

$$\hat{x} = \lim_{M\to\infty} \frac{1}{M}\sum_{s=1}^{M} x_s$$

は等しくなる。

$$\overline{x} = \hat{x}$$

このような時間平均とアンサンブル平均が等しい信号をエルゴード的である，という。

これは，物理学の統計力学で多数の分子の集団の運動について使われる言葉である。分子の運動が熱運動のような規則性を持たないランダム（無秩序）な運動の場合，１つの分子の状態の時間変化は，時間を止めて見た多数の分子の状態の違いと同じである。 このとき，分子の集団の運動はエルゴード的であるという。

2　自己相関関数

自己相関関数は，次のような量 $A(t)$ の時間平均をとったものである。時間的に変化する信号 $v(t)$ について，ある時刻 t での値と，その τ だけ後の時刻での値の積を $A(t)$ とする。

$$A(t) = v(t) \cdot v(t+\tau)$$

もし $v(t)$ が周期性をもっている場合は，$v(t)$ と $v(t+\tau)$ の値には一定の関係があるから積 $v(t) \cdot v(t+\tau)$ の時間平均は τ により値は異なるが，0 でない値になる。一方，$v(t)$ が 0 を中心とする正負の完全にランダムな（無秩序な）信号の場合は，積 $v(t) \cdot v(t+\tau)$ も 0 を中心とする正負のランダムな値になるから，その時間平均は 0 となる。$A(t)$ の時間平均は前節の式（4）を参考にして

$$\overline{A(t)} = \lim_{T\to\infty} \frac{1}{T} \int_0^T A(t)\mathrm{d}t \tag{5}$$

のように表される。$\overline{A(t)}$ は自己相関関数である。次節でもう一度きちんと定義する。

2.1　定義

時間的に変化する信号波形 $v(t)$ についての自己相関関数 $R(\tau)$ を次式で定義する。

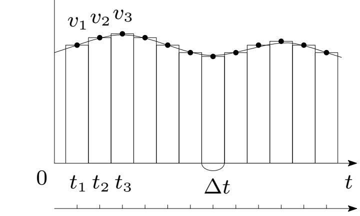
$$R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T v(t) \cdot v(t+\tau)\mathrm{d}t \tag{6}$$

$R(\tau)$ は信号波形について，時間 τ だけ離れた点の相関の大きさを表す量である。τ は遅れ時間と呼ばれるが，積分に関して τ は定数であることに注意する。

2.2　実際のデータにおける自己相関関数の計算

実際の信号処理では，信号は時間 T の間にサンプリングされた有限な個数の時系列データ $v(i)$ として与えられる。この場合には式（6）の積分はサンプリングの時間間隔についての和に置き換えて計算する。

時間的に変化する電圧や電流（信号）の値を時刻 t の関数として図に表したものを**波形**という。信号波形ということもある。いま，サンプリングの時間間隔 Δt で，$v(1), v(2), \cdots, v(i), \cdots, v(n)$ の n 個のデータを得たとする。$v(i)$ は波形の値である。このようなデータを離散的な（連続でない）時系列データと呼ぶ。（前節で v_i と書いたのと同じものを，ここではコンピュータプログラミングの配列変数を意識して $v(i)$ と書く。）



時系列データについて，i 番目の時刻と j 区間だけ後の $i+j$ 番目の時刻をとり，2 つの時刻でのデータ値の積

$$A(i) = v(i) \cdot v(i+j)$$

を計算する。 i は時刻 t に相当し，j はおくれ時間 τ に相当する。 いま任意のデータ数 N 個の測定は，時間 $T = N\Delta t$ で行われ，この間の $A(i)$ の時間平均は

$$\overline{A(i)} = \frac{1}{T}\sum_{i=1}^{N} A(i)\cdot\Delta t = \frac{1}{N}\sum_{i=1}^{N} A(i) = \frac{1}{N}\sum_{i=1}^{N} v(i)\cdot v(i+j)$$

と表される。 これは次節で定義する自己相関関数である。

定義

離散的な時系列データの i 番目と $i+j$ 番目のデータの積 $v(i)\cdot v(i+j)$ の，N 個のデータの測定時間 $T = N\Delta t$ の間における時間平均を自己相関関数という。

$$R(j) = \frac{1}{N\cdot\Delta t}\sum_{i=1}^{N} v(i)\cdot v(i+j)\cdot\Delta t$$

Δt は消えて**自己相関関数**は

$$\boldsymbol{R(j) = \frac{1}{N}\sum_{i=1}^{N} v(i)\cdot v(i+j) \qquad j = 0, 1, 2, \cdots, N} \tag{7}$$

である。 上式で，Σ に関して j は定数であることに注意する。 j として $j = 0, 1, 2, \cdots, N$ の 1 つを与えて式（7）を計算し 1 つの $R(j)$ を得る。

$j = 0$ は特別な場合であり

$$R(0) = \frac{1}{N}\sum_{i=1}^{N} v^2(i) \tag{8}$$

は N 個のデータの 2 乗の平均値で，全ての $R(i)$ の中で最大の値である。これは同じデータ同士の相関を表す量であり，完全に相関がある場合であるから 1 に規格化する。 そのためには全ての $R(i)$ を $R(0)$ で割った値を改めて $R(i)$ とすればよい。 規格化する場合には式（7）の $1/N$ はキャンセルするので初めから省略できる。

データ数

式（7）の計算を行うために全データ数 n は，N の 2 倍を必要とする。なぜなら，$i = N$, $j = N$ のとき，$v(N+N) = v(2N)$ となるから，$v(1) \sim v(n)$ の $n = 2N$ 個のデータが必要である。

2.3　c 言語による自己相関関数プログラミング

次のプログラムは自己相関関数計算のメインの部分である。ファイル autocorr.c としてコンパイル・実行可能であるがこれだけでは何も結果は得られない。 データ入力と出力部分が別に必要である。

プログラムの要点

○時系列データ $v(i)$ $i = 1,\ 2,\ \cdots, n$ を配列変数 `v[i]` に用意する（この部分は以下のプログラムには含まれていない)。 データの個数は $n = 2N$ 個必要である。

○`#define Z 1048676` で `Z`（配列変数の個数）を $1048676 = 2^{20}$（最大）に設定。

○ `float v[Z], r[Z];` で配列変数 `v[Z], r[Z]` を使用宣言。 `v[i]` について自己相関関数の式（7）を計算し，`r[i]` に書出す。

ただし，式（8）のところで述べたように $R(0) = 1$ として全体を規格化するため，最後の行で `r[i]` の全てを `r[0]` で割っている。 従って，式（7）の $1/N$ は不要で，`r[i] = s / N;+` ではなく，`r[i] = s;` としている。

○自己相関関数の計算ではデータ数は 2 の冪乗である必要はないが，FFT（高速フーリエ変換）と連携処理をする場合のため，データ数を 2 の冪乗の個数にする。（FFT 計算ではデータ数は 2 の冪乗であることが必要)。

○ `main{ }` で `autocorr{ }` を呼ぶ。

○　自己相関関数の計算結果は配列変数 `r[j] j=0, 1, 2, … N` に収める。

```
//*****************************************************

//                 autocorr.c

#include <stdio.h>
#define Z 1048676
int i, j, N, n;
float v[Z], r[Z];
void autocorr(float*, float*);//  プロトコル宣言

void main(){
      autocorr(v, r);
}

//  自己相関関数 r[0] ～ r[N] 計算
void autocorr(d, e) float d[ ]; float e[ ];{
      float s;      // Σ用
      N = n/2;
      for (j = 0; j <= N; j++) {
           s = 0;
           for (i = 1; i <= N; i++) {
                s = s + v[i] * v[i+j];
           }
           r[j] = s;
      }
      if(r[0]) for (j = 0; j <= N; j++) r[j] = r[j] / r[0];
}
//*****************************************************
```