高知工科大学

基礎数学ワークブック

(2001年度版)

番外編 2

「データの関数近似」

内容

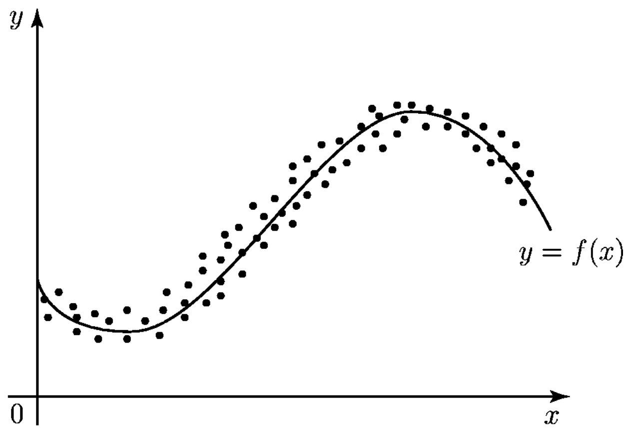
◎ 補間

◎ 対数方眼紙

◎ フーリエ近似

◎ 回帰直線

◎ 最小2乗法



電子・光システム工学科

井上 昌昭 著

< この号の構成 >

この号では主に 2 変数のデータ $(x_0, y_0), (x_1, y_1), \cdots, (x_n, y_n)$ に対し、x と y の関係を求める方法を解説する。

< I. 関数近似 >

この号の前半は x と y の関係が関数 $y = f(x)$ として表される場合を考える。

(1) $y = f(x)$ が x の多項式になっている場合
　$\Rightarrow$ ラグランジュの補間多項式より $f(x)$ の形を決める。 (P2～P4)

(2) $y = f(x)$ が x の指数関数になっている場合
　$\Rightarrow$ 片対数方眼紙によって指数関数の式を求める。 (P5 , P6)

(3) $y = f(x)$ が x の巾乗 $y = x^{\square}$ の形になっている場合
　$\Rightarrow$ 両対数方眼紙によって巾乗 $x^{\square}$ の式を求める。 (P7 , P8)

(4) $y = f(x)$ が x の三角多項式になっている場合
　$\Rightarrow$ 三角多項式補間により $f(x)$ の形を決める。 (P9～P17)

(4)′ 周期関数 $f(x)$ が与えられたとき、三角多項式で近似するのに必要なデータ数はフーリエ近似の例より算出する。 (P18 , P19)

(注) 離散フーリエ変換 (P20～P22) はフーリエ近似の複素数表示である。

< II. 2変数データの関係 >

この号の後半は x と y との関係がはっきりわからない場合に、その傾向を判断する方法を考える。

(1) x と y のだいたいの傾向を調べるためには、まず散布図を作り、相関係数 r_{xy} を計算する。 (P23 , P28)

(2) y が x の一次式で近似できる場合
　$\Rightarrow$ 回帰直線を求める。 (P24～P27)

(3) 回帰直線のあてはまりの尺度は決定係数を計算する。 (P30)

(4) y が x の多項式で近似できる場合
　$\Rightarrow$ 回帰多項式を求める (P35) 。数学的には最小 2 乗法 (P31～P34) を用いる。

(5) x と y をまとめて 1 つの変数 $z = ax + by$ にしたいときは、直交回帰直線を計算する (P37～P40) 。

(注) 3 変数以上のデータ $x, y, z, \cdots$ を一次式 $ax + by + cz + \cdots$ で近似したい場合は重回帰式を用いる (P36) 。

＜ラグランジュ補間 1＞

例 1　(2 点を通る直線)

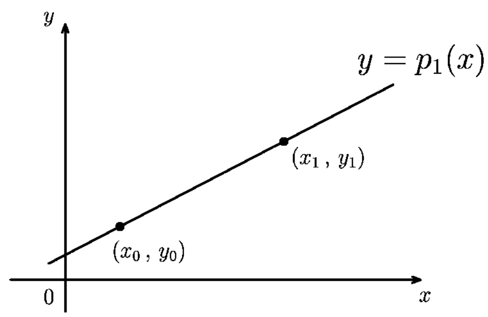
2 点 $(x_0\,,\,y_0)\,,(x_1\,,\,y_1)$ を通る直線の方程式は

$$y-y_0=\frac{y_1-y_0}{x_1-x_0}(x-x_0)\quad\cdots\cdots\cdots\cdots\quad(1)$$

である。これを式変形すると

$$y=\frac{x-x_1}{x_0-x_1}\times y_0+\frac{x-x_0}{x_1-x_0}\times y_1\quad\cdots\quad(2)$$

となる。



問 1　(1) 式を変形して (2) 式を導け。

問 2　$p_1(x)=\dfrac{x-x_1}{x_0-x_1}\times y_0+\dfrac{x-x_0}{x_1-x_0}\times y_1$ とおく。$p_1(x_0)$ と $p_1(x_1)$ の値を求めよ。

例 2　(3 点を通る放物線)

3 点 $(x_0\,,\,y_0)\,,(x_1\,,\,y_1)\,,(x_2\,,\,y_2)$ を通る放物線はただ 1 つである。放物線の式を

$$y=ax^2+bx+c$$

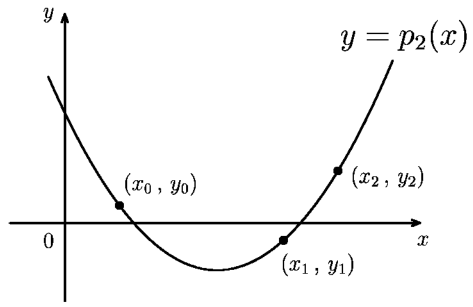
とおくと $a,\,b,\,c$ に関する連立方程式

$$\begin{cases} y_0=ax_0^2+bx_0+c \\ y_1=ax_1^2+bx_1+c \\ y_2=ax_2^2+bx_2+c \end{cases}$$

を解けばよい。この連立方程式を解くかわりに

$$p_2(x)=\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}\times y_0+\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}\times y_1+\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}\times y_2$$

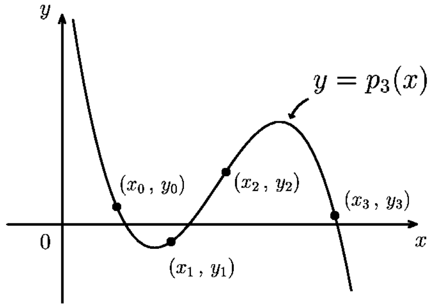
とおくと $y=p_2(x)$ が求める放物線の式である。



問 3　$p_2(x_0)\,,\,p_2(x_1)\,,\,p_2(x_2)$ の値を求めよ。

問 4　3 点 $(1,\,1),\,(3,\,-1),\,(4,\,4)$ を通る放物線の方程式を求めよ。

＜ラグランジュ補間 2＞



例 1　(4 点を通る 3 次曲線)

4 点 $(x_0\,,\,y_0)\,,\,(x_1\,,\,y_1)\,,\,(x_2\,,\,y_2)\,,\,(x_3\,,\,y_3)$

を通る 3 次関数は一意的に定まる。

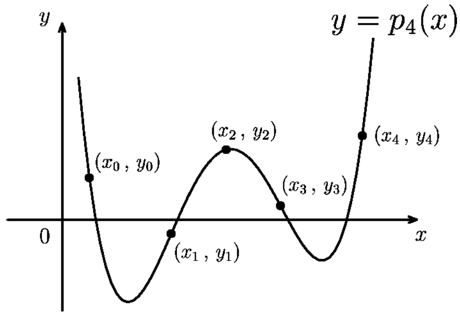
上の 4 点に対し

$$p_3(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} \times y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} \times y_1$$
$$+\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} \times y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} \times y_3$$

とおくと $p_3(x)$ は x の 3 次関数である。

問 1　上の関数 $p_3(x)$ に対し、$x = x_0\,,\,x = x_1\,,\,x = x_2\,,\,x = x_3$ のときの値を求めよ。

$p_3(x_0) =$ 　　　　　　, $p_3(x_1) =$ 　　　　　　, $p_3(x_2) =$ 　　　　　　, $p_3(x_3) =$



例 2　(5 点を通る 4 次曲線)

5 点 $(x_0\,,\,y_0)\,,\,(x_1\,,\,y_1)\,,\,(x_2\,,\,y_2)\,,\,(x_3\,,\,y_3)\,,\,(x_4\,,\,y_4)$

を通る 4 次関数は一意的に定まる。

上の 5 点に対し

$$p_4(x) = \frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)} \times y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)} \times y_1$$
$$+\frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)} \times y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)(x_3-x_4)} \times y_3$$
$$+\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_0)(x_4-x_1)(x_4-x_2)(x_4-x_3)} \times y_4$$

とおくと $p_4(x)$ は x の 4 次関数である。

問 2　上の $p_4(x)$ に対し、$x = x_0\,,\,x = x_1\,,\,x = x_2\,,\,x = x_3\,,\,x = x_4$ のときの値を求めよ。

$p_4(x_0) =$ 　　　　, $p_4(x_1) =$ 　　　　, $p_4(x_2) =$ 　　　　, $p_4(x_3) =$ 　　　　, $p_4(x_4) =$

＜ラグランジュ補間 3 ＞

$a_1+a_2+a_3+a_4+a_5=\sum_{k=1}^{5} a_k$ のように和を記号 $\sum$ を用いて表す。

これと同様にして積 (Product) をアルファベットの P に相当するギリシャ文字 $\prod$ を用いて

$$a_1\times a_2\times a_3\times a_4\times a_5=\prod_{i=1}^{5} a_i \quad \left(\text{または} \prod_{1\leqq i\leqq 5} a_i\right)$$

のように表す。また途中の項がない積を

$$a_1\times a_2\times a_4\times a_5=\prod_{\substack{1\leqq i\leqq 5\\ i\neq 3}} a_i$$

のように表す。

例　前ページ例 1 の $p_3(x)$ は記号 $\prod$ を使うと

$$p_3(x)=\frac{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 0}}(x-x_i)}{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 0}}(x_0-x_i)}\times y_0+\frac{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 1}}(x-x_i)}{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 1}}(x_1-x_i)}\times y_1+\frac{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 2}}(x-x_i)}{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 2}}(x_2-x_i)}\times y_2+\frac{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 3}}(x-x_i)}{\prod_{\substack{0\leqq i\leqq 3\\ i\neq 3}}(x_3-x_i)}\times y_3$$

と表される。さらに和の記号 $\sum$ を用いると

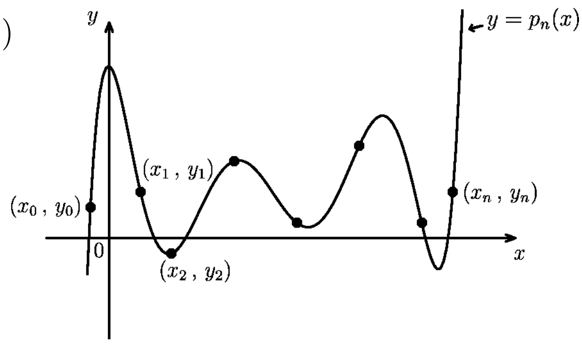
$$p_3(x)=\sum_{k=0}^{3}\frac{\prod_{\substack{0\leqq i\leqq 3\\ i\neq k}}(x-x_i)}{\prod_{\substack{0\leqq i\leqq 3\\ i\neq k}}(x_k-x_i)}\times y_k$$

と表される。

問　前ページ例 2 の $p_4(x)$ を記号 $\prod$ と $\sum$ を用いて表せ。

一般に $n+1$ 個の点 $(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$ を通る n 次関数を $p_n(x)$ とすると、

$$\boxed{p_n(x)=\sum_{k=0}^{n}\frac{\prod_{\substack{0\leqq i\leqq n\\ i\neq k}}(x-x_i)}{\prod_{\substack{0\leqq i\leqq n\\ i\neq k}}(x_k-x_i)}\times y_k}$$



となる。この $p_n(x)$ をラグランジュ(Lagrange) の**補間多項式**という。

< 片対数方眼紙 1 >

常用対数 (底が 10 の対数)

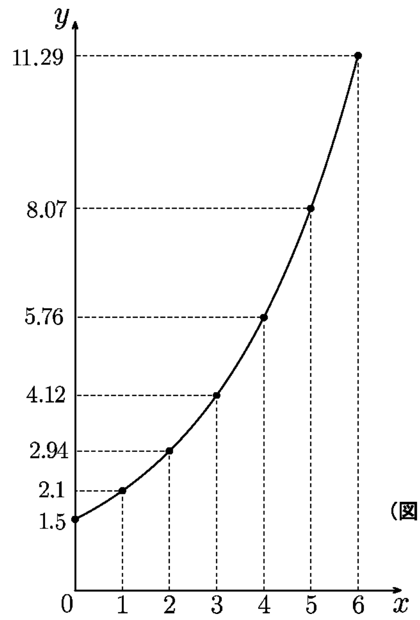
$$0 = \log_{10} 1,\quad 1 = \log_{10} 10,\quad 2 = \log_{10} 100,\quad 3 = \log_{10} 1000,\quad 4 = \log_{10} 10000$$

で目盛った目盛を **対数目盛** という。縦または横の一方の目盛を対数目盛とした方眼紙を **片対数方眼紙** または **半対数方眼紙** という。

例 表 1 のような x と y の数値データを xy 平面上にプロットしたグラフが図 1 である。

x	0	1	2	3	4	5	6
y	1.5	2.1	2.94	4.12	5.76	8.07	11.29

(表 1)



(図 1)

x と y の関係を求めたい。
このデータを片対数方眼紙 (図 2) にプロットすると直線になる。この直線の傾きは

$$\text{傾き} = \frac{\log_{10}(11.29) - \log_{10}(1.5)}{6-0} = \frac{1}{6}\log_{10}\left(\frac{11.29}{1.5}\right)$$
$$\fallingdotseq 0.146$$

であるから Y と x の関係は

$$Y = 0.146x + \log_{10} 1.5$$

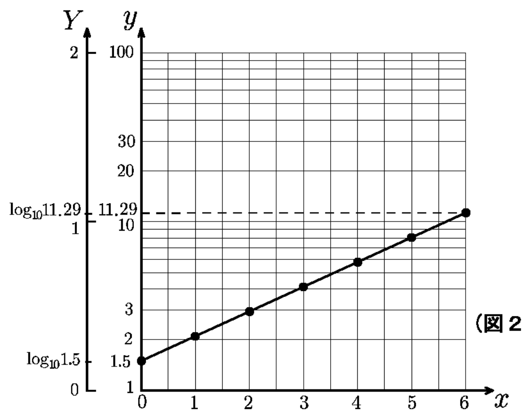
となる。ここで Y と y には

$$\boxed{Y = \log_{10} y}$$

の関係があるから、y と x の関係は

$$\log_{10} y = 0.146x + \log_{10} 1.5$$
$$\Downarrow$$
$$y = 10^{0.146x + \log_{10} 1.5} \Rightarrow y = 10^{0.146x} \times 10^{\log_{10} 1.5} \Rightarrow \boldsymbol{y = 1.5 \times 10^{0.146x}}$$

となる。

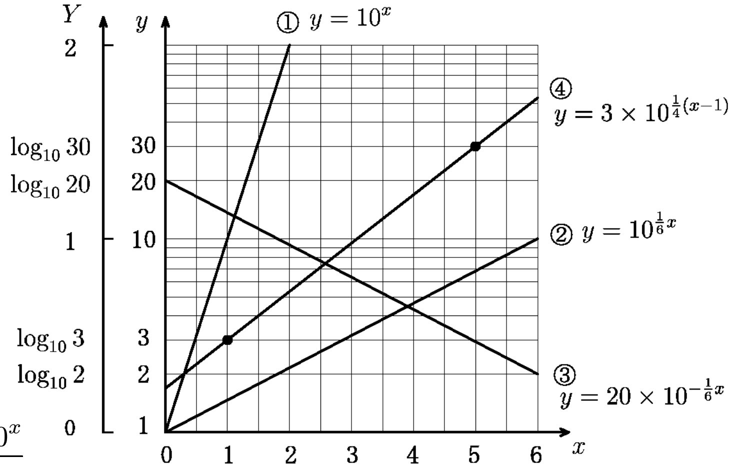


(図 2)

(注) 対数の性質より

$$A = \log_{10} 1.5 \iff 10^A = 1.5 \iff 10^{\log_{10} 1.5} = 1.5$$

＜片対数方眼紙2＞



例　右図のような片対数方眼紙では対数目盛 y と標準目盛 Y の間に

$$\boxed{Y=\log_{10}y} \quad (*)$$

の関係がある。右の 4 本の直線が表す関数 (y と x の関係式) を求める。

(1) 直線①は

$Y=x\Rightarrow\log_{10}y=x\Rightarrow\underline{y=10^x}$

(2) 直線②は

$$Y=\frac{1}{6}x\ \Rightarrow\ \log_{10}y=\frac{1}{6}x\ \Rightarrow\ \underline{y=10^{\frac{1}{6}x}}$$

(3) 直線③は

$$傾き=\frac{\log_{10}2-\log_{10}20}{6-0}=-\frac{1}{6}(\log_{10}20-\log_{10}2)=-\frac{1}{6}\log_{10}\left(\frac{20}{2}\right)=-\frac{1}{6}$$

より

$$Y=-\frac{1}{6}x+\log_{10}20\ \Rightarrow\ \log_{10}y=-\frac{1}{6}x+\log_{10}20\ \Rightarrow\ \log_{10}y=\log_{10}\left(10^{-\frac{1}{6}x}\times20\right)\ \Rightarrow\ \underline{y=20\times10^{-\frac{1}{6}x}}$$

(4) 直線④は

$x=1$ のとき $Y=\log_{10}3$,　$x=5$ のとき $Y=\log_{10}30$

より

$$傾き=\frac{\log_{10}30-\log_{10}3}{5-1}=\frac{1}{4}\log_{10}\left(\frac{30}{3}\right)=\frac{1}{4}$$

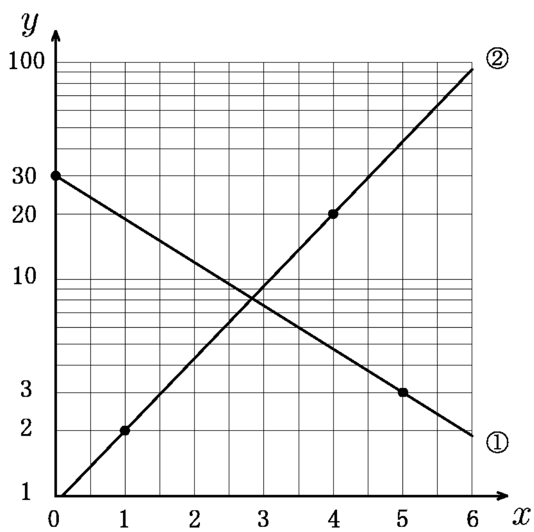
だから

$$Y=\frac{1}{4}(x-1)+\log_{10}3\ \Rightarrow\ y=10^{\frac{1}{4}(x-1)+\log_{10}3}\ \Rightarrow\ \underline{y=3\times10^{\frac{1}{4}(x-1)}}$$

(注) 通常の片対数方眼紙には標準目盛 Y がない。

問 1　右図の 2 本の直線①、②が表す関数 (y と x の関係式) を求めよ。

①　　　　　　　　②



問 2　自然対数の底 $e\fallingdotseq2.718$(ネピアの数) に対し、$\log_e10\fallingdotseq2.3$

より

$$10=e^{\log_e10}\fallingdotseq e^{2.3}$$

だから例の②は

$$10^{\frac{1}{6}x}=\left(e^{2.3}\right)^{\frac{1}{6}x}=e^{\frac{2.3}{6}x}$$

となる。問 1 で求めた関数を $y=\square\times e^{\square}$ の形にせよ。

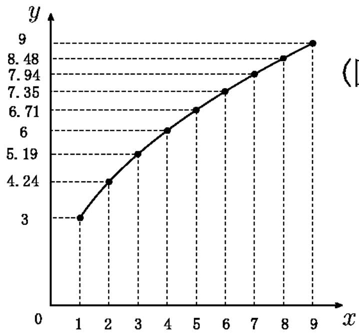
＜両対数方眼紙１＞

縦軸と横軸の両方が対数目盛りの方眼紙を**両対数方眼紙**または**全対数方眼紙**という。

例 表 1 のような x と y の数値データを xy 平面上にプロットしたグラフが図 1 である。

x	1	2	3	4	5	6	7	8	9
y	3.0	4.24	5.19	6.0	6.71	7.35	7.94	8.48	9.0

(表 1)



(図1)

x と y の関係を求めたい。

このデータを両対数方眼紙 (図 2) にプロットすると直線になる。

この直線を標準目盛 X,Y でみると

$X = \log_{10} 4$ のとき $Y = \log_{10} 6$

$X = 0$ のとき $Y = \log_{10} 3$

より傾きは

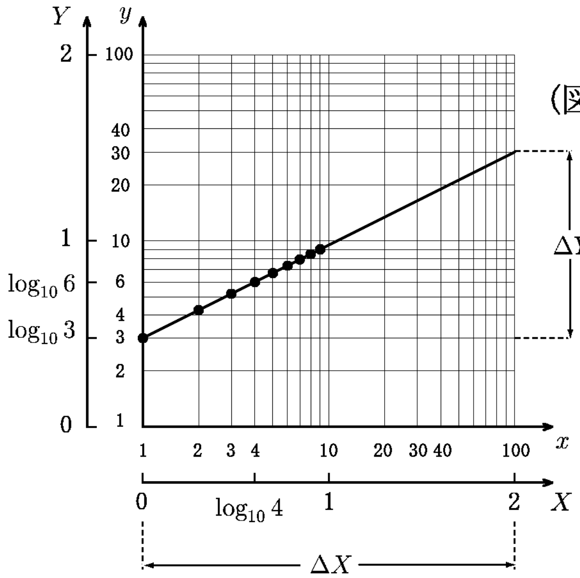
$$傾き = \frac{\log_{10} 6 - \log_{10} 3}{\log_{10} 4 - 0} = \frac{1}{2}$$

であるから X,Y の関係は

$$Y = \frac{1}{2}X + \log_{10} 3 \quad \cdots (1)$$

となる。一方 X,Y と x,y の間には

$$\boxed{X = \log_{10} x,\ \ Y = \log_{10} y}$$



(図2)

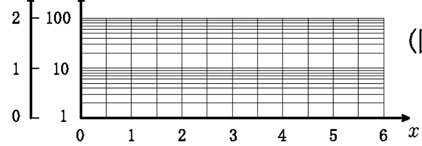
の関係があるので (1) 式より

$$\log_{10} y = \frac{1}{2}\log_{10} x + \log_{10} 3 = \log_{10}\left(x^{\frac{1}{2}} \times 3\right) = \log_{10}(3\sqrt{x})$$

よって求める x と y の関係式は $\underline{y = 3\sqrt{x}}$ である。

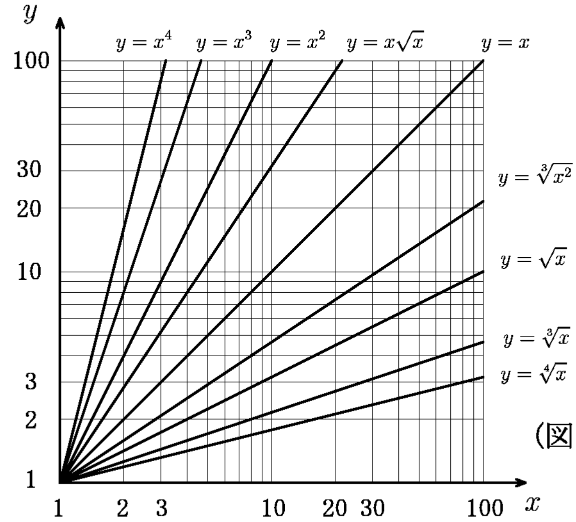
(注 1) 例の場合は直線の傾きを計算で求めたが、図 2 の ΔX と ΔY の長さを実際にものさしで計って、その比 $\dfrac{\Delta Y}{\Delta X}$ をもとめればよい。傾き $= \dfrac{\Delta Y}{\Delta X} = \dfrac{1}{2}$ が求まる。両対数方眼紙の場合は (X と Y の目盛幅が同じなので) 実際に傾きを計ればよい。

(注 2) 片対数方眼紙の場合、直線の傾きは前ページのように計算で求めるしかない。x と y の目盛幅が違うからである。もし同じにしたら図 3 のようになって点をプロットしにくい。



(図3)

＜ 両対数方眼紙2 ＞



(図1)

例1　x の巾乗 $y = x^{\square}$ のグラフを両対数方眼紙に書くと、図 1 のような直線になる。

$y = x^2$　⇒　傾き 2 の直線

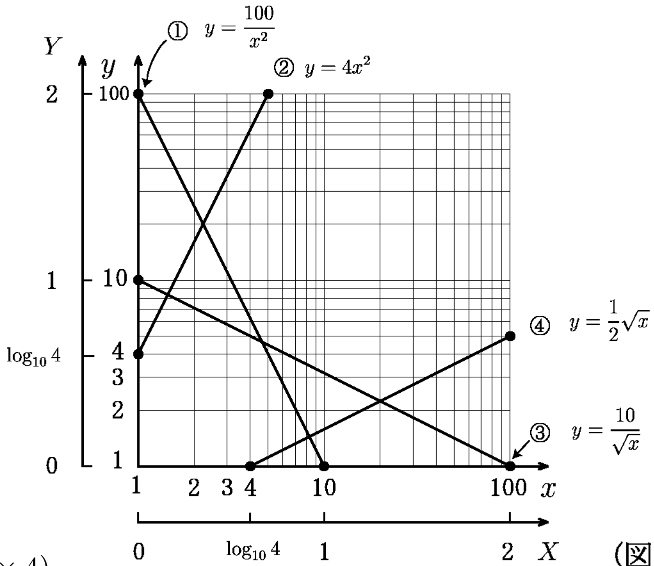
$y = \sqrt{x} = x^{\frac{1}{2}}$　⇒　傾き $\frac{1}{2}$ の直線

$y = x\sqrt{x} = x^{\frac{3}{2}}$　⇒　傾き $\frac{3}{2}$ の直線

例2　図 2 の 4 本の直線が表す関数 (x と y の関係式) を求めたい。
理解しやすいために標準目盛 X, Y を用意する。対数目盛 x, y との間には

$$\boxed{X = \log_{10} x, \quad Y = \log_{10} y}$$

の関係がある。



(図2)

(1) 直線①は

$Y = -2X + \log_{10} 100$

$\Rightarrow \log_{10} y = -2\log_{10} x + \log_{10} 100$

$\Rightarrow \underline{y = x^{-2} \times 100 = \frac{100}{x^2}}$

(2) 直線②は

$$Y = \left(\frac{\log_{10} 100 - \log_{10} 4}{\log_{10} 5}\right) X + \log_{10} 4$$

$= 2X + \log_{10} 4$

$\Rightarrow \log_{10} y = 2\log_{10} x + \log_{10} 4 = \log_{10}(x^2 \times 4)$

$\Rightarrow \underline{y = 4x^2}$

(3) 直線③は

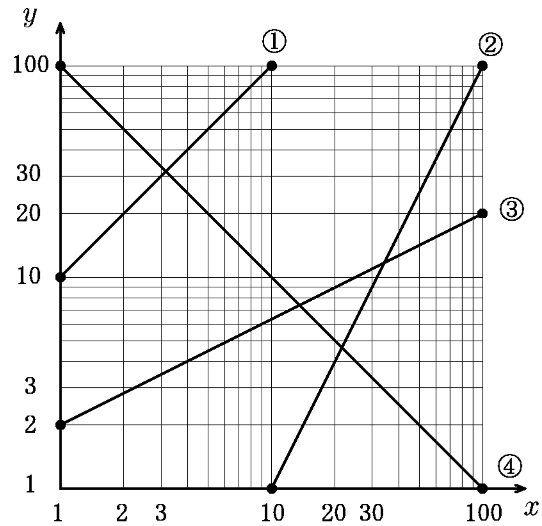
$Y = -\frac{1}{2}X + \log_{10} 10 \Rightarrow \underline{y = x^{-\frac{1}{2}} \times 10 = \frac{10}{\sqrt{x}}}$

(4) 直線④は

$$Y = \left(\frac{\log_{10} 5}{\log_{10} 100 - \log_{10} 4}\right)(X - \log_{10} 4)$$

$= \frac{1}{2}X - \log_{10} 2$

$\Rightarrow \underline{y = x^{\frac{1}{2}} \div 2 = \frac{1}{2}\sqrt{x}}$



(図3)

問　図 3 の 4 本の直線が表す関数 (x と y の関係式) を求めよ。

①　　　　　　②

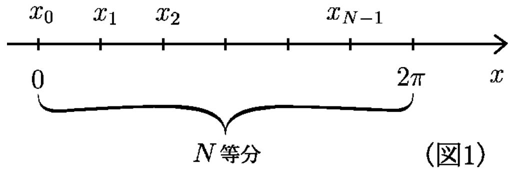
③　　　　　　④

＜三角多項式補間１＞

N 個の 2 変数データ

$$(x_0, y_0), (x_1, y_1), \cdots, (x_{N-1}, y_{N-1})$$

がある。図 1 のように $x_0, x_1, \cdots, x_{N-1}$ が 0 から 2π までを N 等分した分点



(図1)

$$x_0 = 0\ ,\quad x_1 = \frac{2\pi}{N}\ ,\quad \cdots\ ,\quad x_l = \frac{2\pi l}{N}\ ,\quad \cdots\ ,\quad x_{N-1} = \frac{2\pi(N-1)}{N}$$

である場合には以下のような三角多項式による補間公式がなりたつ。

〔I〕＜ $N = 2n-1$ (奇数) の場合 ＞

(1) $$\boxed{f_{2n-1}(x) = A_0 + 2\sum_{k=1}^{n-1}\Big\{A_k\cos(kx) + B_k\sin(kx)\Big\}}$$

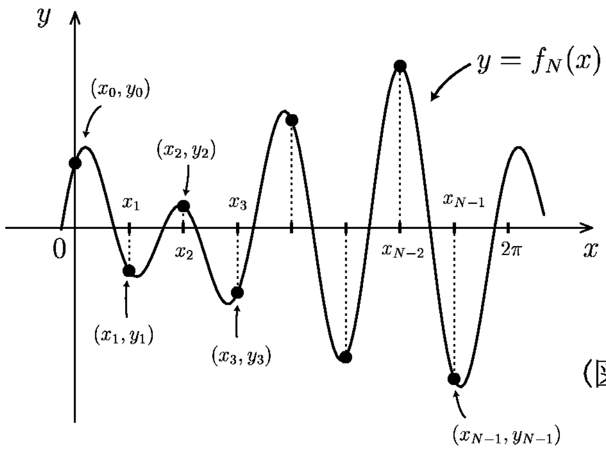
の形の三角多項式に対し、係数が

(∗) $$\boxed{A_k = \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell\cos(kx_\ell)\ ,\quad B_k = \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell\sin(kx_\ell)}$$

であれば $y = f_{2n-1}(x)$ のグラフは N 個の点

$$(x_0, y_0), (x_1, y_1), \cdots, (x_{N-1}, y_{N-1})$$

を通る曲線になる。(図 2)



(図2)

〔II〕＜ $N = 2n$ (偶数) の場合 ＞

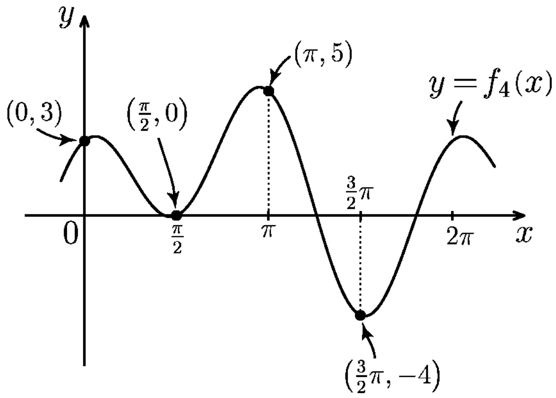
(2) $$\boxed{f_{2n}(x) = A_0 + 2\sum_{l=0}^{n-1}\Big\{A_k\cos(kx) + B_k\sin(kx)\Big\} + A_n\cos(nx)}$$

の形の三角多項式に対し、各係数が上と同じ (∗) 式を満たせば、

$y = f_{2n}(x)$ のグラフは N 個の点 $(x_0, y_0), (x_1, y_1), \cdots, (x_{N-1}, y_{N-1})$ を通る曲線になる。

＜三角多項式補間 2＞

前ページの三角多項式による補間公式の使い方を例をあげて説明する。応用上は $N=2n$(偶数) の場合がよく使われるので、N が偶数の例をあげる。



例 右図の 4 点

$(0,3)$, $\left(\frac{\pi}{2},0\right)$ $(\pi,5)$ $\left(\frac{3}{2}\pi,-4\right)$

を通る三角多項式 $f_4(x)$ の式を求めたい。この場合は $N=4$,

$x_0=0$, $x_1=\frac{\pi}{2}$, $x_2=\pi$, $x_3=\frac{3}{2}\pi$,

$y_0=3$, $y_1=0$, $y_2=5$, $y_3=-4$

である。前ページ $(*)$ 式より

$$A_0=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\cos(0)=\frac{1}{4}\sum_{\ell=0}^{3}y_\ell=\frac{1}{4}(y_0+y_1+y_2+y_3)=\frac{1}{4}(3+0+5-4)=1$$

$$A_1=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\cos(x_\ell)=\frac{1}{4}\left\{3\times\cos(0)+0\times\cos\left(\frac{\pi}{2}\right)+5\times\cos(\pi)-4\times\cos\left(\frac{3}{2}\pi\right)\right\}=-\frac{1}{2}$$

$$A_2=\frac{1}{N}\sum_{2\ell=0}^{N-1}y_\ell\cos(2x_\ell)=\frac{1}{4}\left\{3\times\cos(0)+0\times\cos(\pi)+5\times\cos(2\pi)-4\times\cos(3\pi)\right\}=3$$

$$B_1=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\sin(x_\ell)=\frac{1}{4}\left\{3\times\sin(0)+0\times\sin\left(\frac{\pi}{2}\right)+5\times\sin(\pi)-4\times\sin\left(\frac{3}{2}\pi\right)\right\}=1$$

である。前ページ (2) 式より $n=2$ であるから

$$\begin{aligned}f_4(x)&=A_0+2\sum_{k=1}^{2-1}\left\{A_k\cos(kx)+B_k\sin(kx)\right\}+A_n\cos(nx)\\&=A_0+2\left\{A_1\cos x+B_1\sin x\right\}+A_2\cos(2x)\\&=1-\cos x+2\sin x+3\cos(2x)\end{aligned}$$

となる。

問 $f_4(x)=1-\cos x+2\sin x+3\cos(2\pi)$ に対し、次の関数の値を求めよ。

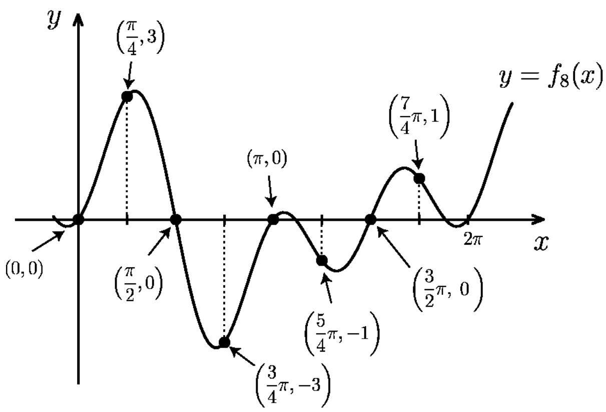
(1) $f_4(0)=$

(2) $f_4\left(\frac{\pi}{2}\right)=$

(3) $f_4(\pi)=$

(4) $f_4\left(\frac{3}{2}\pi\right)=$

＜三角多項式補間3＞



例 右図の 8 点を通る三角多項式 $f_8(x)$ の式を求めたい。8 点の座標 (x_ℓ, y_ℓ) は以下の通り。

ℓ	0	1	2	3	4	5	6	7
x_ℓ	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3}{4}\pi$	π	$\frac{5}{4}\pi$	$\frac{3}{2}\pi$	$\frac{7}{4}\pi$
y_ℓ	0	3	0	-3	0	-1	0	1

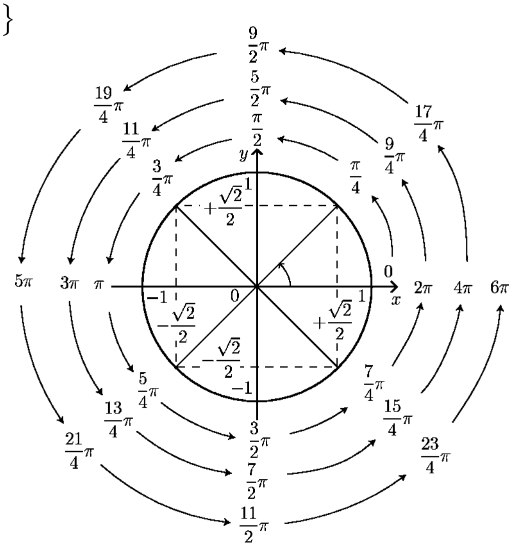
9 ページの (2) 式より

$$f_8(x) = A_0 + 2\sum_{k=1}^{3}\left\{A_k\cos(kx) + B_k\sin(kx)\right\} + A_4\cos(4x)$$

となる。各係数 A_0, A_1, A_2, A_3, A_4, B_1, B_2, B_3 を求めたい。

$$A_0 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\cos 0 = \frac{1}{8}\left\{y_0 + y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7\right\}$$

$$= \frac{1}{8}\left\{0 + 3 + 0 - 3 + 0 - 1 + 0 + 1\right\} = 0$$



$$A_1 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\cos(x_\ell)$$

$$= \frac{1}{8}\left\{0\times\cos 0 + 3\times\cos\left(\frac{\pi}{4}\right) + 0\times\cos\left(\frac{\pi}{2}\right) - 3\times\cos\left(\frac{3}{4}\pi\right)\right.$$

$$\left.+0\times\cos\pi - 1\times\cos\left(\frac{5}{4}\pi\right) + 0\times\cos\left(\frac{3}{2}\pi\right) + 1\times\cos\left(\frac{7}{4}\pi\right)\right\}$$

$$= \frac{1}{8}\left\{0 + \frac{3\sqrt{2}}{2} + 0 - \left(-\frac{3\sqrt{2}}{2}\right) + 0 - \left(-\frac{\sqrt{2}}{2}\right) + 0 + \frac{\sqrt{2}}{2}\right\} = \frac{\sqrt{2}}{2}$$

$$A_3 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\cos(3x_\ell)$$

$$= \frac{1}{8}\left\{0\times\cos 0 + 3\times\cos\left(\frac{3}{4}\pi\right) + 0\times\cos\left(\frac{3}{2}\pi\right) - 3\times\cos\left(\frac{9}{4}\pi\right)\right.$$

$$\left.+0\times\cos(3\pi) - 1\times\cos\left(\frac{15}{4}\pi\right) + 0\times\cos\left(\frac{9}{2}\pi\right) + 1\times\cos\left(\frac{21}{4}\pi\right)\right\}$$

$$= \frac{1}{8}\left\{0 + 3\times\left(-\frac{\sqrt{2}}{2}\right) + 0 - 3\times\frac{\sqrt{2}}{2} + 0 - \frac{\sqrt{2}}{2} + 0 + \left(-\frac{\sqrt{2}}{2}\right)\right\} = -\frac{\sqrt{2}}{2}$$

問1 残りの係数 A_2, A_4, B_1, B_2, B_3 の値を求めよ。

$$A_2 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\cos(2x_\ell) =$$

$$A_4 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\cos(4x_\ell) =$$

$$B_1 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\sin(x_\ell) =$$

$$B_2 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\sin(2x_\ell) =$$

$$B_3 = \frac{1}{8}\sum_{\ell=0}^{7} y_\ell\sin(3x_\ell) =$$

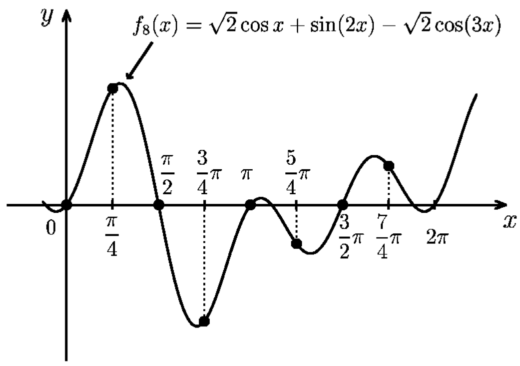
＜三角多項式補間 4 ＞

例　前ページの例の場合には

$$A_0 = 0\,,\ A_1 = \frac{\sqrt{2}}{2}\,,\ A_2 = 0\,,\ A_3 = -\frac{\sqrt{2}}{2}$$

$$A_4 = 0\,,\ B_1 = 0\,,\ B_2 = \frac{1}{2}\,,\ B_3 = 0$$

であった。この場合の三角多項式 $f_8(x)$ は



$$f_8(x) = A_0 + 2\sum_{k=1}^{3}\{A_k\cos(kx) + B_k\sin(kx)\} + A_4\cos(4x)$$

$$= A_0 + 2\{A_1\cos x + B_1\sin x + A_2\cos(2x) + B_2\sin(2x) + A_3\cos(3x) + B_3\cos(3x)\} + A_4\cos(4x)$$

$$= 0 + 2\left\{\frac{\sqrt{2}}{2}\cos x + 0 + 0 + \frac{1}{2}\sin(2x) - \frac{\sqrt{2}}{2}\cos(3x) + 0\right\} + 0$$

$$= \sqrt{2}\cos x + \sin(2x) - \sqrt{2}\cos(3x)$$

である。前ページの 8 点 $(x_0, y_0), (x_1, y_1), \cdots, (x_7, y_7)$ を通るかどうか確認する。

$f_8(x_0) = f_8(0) = \sqrt{2}\cos 0 + \sin 0 - \sqrt{2}\cos 0 = \sqrt{2} - \sqrt{2} = 0$ より点 $(0, 0)$ を通る。

$$f_8(x_1) = f_8\left(\frac{\pi}{4}\right) = \sqrt{2}\cos\left(\frac{\pi}{4}\right) + \sin\left(\frac{\pi}{2}\right) - \sqrt{2}\cos\left(\frac{3\pi}{4}\right) = \sqrt{2}\times\left(\frac{\sqrt{2}}{2}\right) + 1 - \sqrt{2}\times\left(-\frac{\sqrt{2}}{2}\right) = 1 + 1 + 1 = 3$$

より点 $\left(\frac{\pi}{4}, 3\right)$ を通る。

$f_8(x_2) = f_8\left(\frac{\pi}{2}\right) = \sqrt{2}\cos\left(\frac{\pi}{2}\right) + \sin(\pi) - \sqrt{2}\cos\left(\frac{3}{2}\pi\right) = 0 + 0 - 0 = 0$ より点 $\left(\frac{\pi}{2}, 0\right)$ を通る。

$$f_8(x_3) = f_8\left(\frac{3\pi}{4}\right) = \sqrt{2}\cos\left(\frac{3\pi}{4}\right) + \sin\left(\frac{3\pi}{2}\right) - \sqrt{2}\cos\left(\frac{9\pi}{4}\right) = \sqrt{2}\times\left(-\frac{\sqrt{2}}{2}\right) + (-1) - \sqrt{2}\times\frac{\sqrt{2}}{2} = -1 - 1 - 1$$
$$= -3$$

より点 $\left(\frac{3\pi}{4}, -3\right)$ を通る。

問　$f_8(x) = \sqrt{2}\cos x + \sin(2x) - \sqrt{2}\cos(3x)$ に対し、以下の関数の値を求めよ。

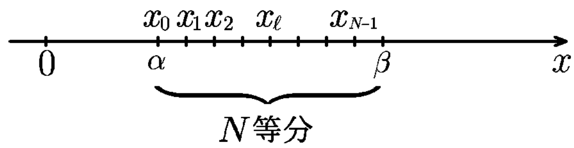
(1)　$f_8(x_4) = f_8(\pi) =$

(2)　$f_8(x_5) = f_8\left(\frac{5\pi}{4}\right) =$

(3)　$f_8(x_6) = f_8\left(\frac{3\pi}{2}\right) =$

(4)　$f_8(x_7) = f_8\left(\frac{7\pi}{4}\right) =$

< 三角多項式補間 5 >



9 ページの公式を一般化する。

$x_0, x_1, \cdots, x_{N-1}$ が区間 $[\alpha, \beta]$ を N 等分した分点である場合、すなわち

$$x_0=\alpha,\ \ x_1=\alpha+\tfrac{\beta-\alpha}{N},\ \ \cdots,\ \ x_\ell=\alpha+\left(\tfrac{\beta-\alpha}{N}\right)\times\ell,\ \ \cdots,\ \ x_{N-1}=\alpha+\left(\tfrac{\beta-\alpha}{N}\right)\times(N-1)$$

である場合には以下のような三角多項式による補間公式が成り立つ。

[I] <$N = 2n-1$(奇数) の場合>

(1) $$\boxed{f_{2n-1}(x)=A_0+2\sum_{k=1}^{n-1}\left\{A_k\cos\left(\tfrac{2\pi k}{\beta-\alpha}(x-\alpha)\right)+B_k\sin\left(\tfrac{2\pi k}{\beta-\alpha}(x-\alpha)\right)\right\}}$$

の形の三角多項式に対し、係数が

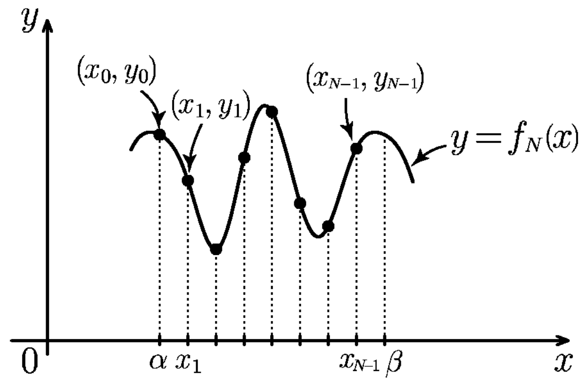
(∗) $$\boxed{A_k=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\cos\left(\frac{2\pi k\ell}{N}\right),\quad B_k=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\sin\left(\frac{2\pi k\ell}{N}\right)}$$

であれば $y=f_{2n-1}$ のグラフは

N 個の点

$(x_0, y_0),\ (x_1, y_1),\ \cdots,\ (x_{N-1}, y_{N-1})$

を通る曲線になる。(右図)



[II] <$N = 2n$(偶数) の場合>

(2) $$\boxed{f_{2n}(x)=A_0+2\sum_{k=1}^{n-1}\left\{A_k\cos\left(\tfrac{2\pi k}{\beta-\alpha}(x-\alpha)\right)+B_k\sin\left(\tfrac{2\pi k}{\beta-\alpha}(x-\alpha)\right)\right\}+A_n\cos\left(\tfrac{2\pi n}{\beta-\alpha}(x-\alpha)\right)}$$

の形の三角多項式に対し、各係数が上と同じ (∗) 式をみたせば、

$y=f_{2n}(x)$ のグラフは N 個の点 $(x_0, y_0),\ (x_1, y_1),\ \cdots,\ (x_{N-1}, y_{N-1})$ を通る曲線になる。

(注) $\alpha=0$, $\beta=2\pi$ の場合は 9 ページと同じ式になる。

＜三角多項式補間6＞

$N=2n$(偶数) の場合に前ページの $f_{2n}(x)$ の係数 A_k , B_k の定義式 $(*)$ を説明する。

$$x_\ell=\alpha+\left(\frac{\beta-\alpha}{N}\right)\ell=\alpha+\left(\frac{\beta-\alpha}{2n}\right)\ell \quad (\ell=0,\ 1,\ \cdots,\ N-1)\text{ より}$$

$$\begin{aligned}f_{2n}(x_\ell)&=A_0+2\sum_{k=1}^{n-1}\left\{A_k\cos\left(\frac{2\pi k}{\beta-\alpha}(x_\ell-\alpha)\right)+B_k\sin\left(\frac{2\pi k}{\beta-\alpha}(x_\ell-\alpha)\right)\right\}+A_n\cos\left(\frac{2\pi n}{\beta-\alpha}(x_\ell-\alpha)\right)\\&=A_0+2\sum_{k=1}^{n-1}\left\{A_k\cos\left(\frac{\pi k\ell}{n}\right)+B_k\sin\left(\frac{\pi k\ell}{n}\right)\right\}+A_n\cos(\pi\ell)\end{aligned}$$

となる。一方三角関数の性質より

$$\boxed{\begin{aligned}&\sum_{\ell=0}^{2n-1}\cos\left(\frac{\pi k\ell}{n}\right)\cos\left(\frac{\pi m\ell}{n}\right)=\begin{cases}0 & : \ k\neq m \quad (0\leqq k,\ m\leqq n)\\ n & : \ k=m \quad (1\leqq k\leqq n-1)\\ 2n & : \ k=m=0 \text{ または } k=m=n\end{cases}\\&\sum_{\ell=0}^{2n-1}\cos\left(\frac{\pi k\ell}{n}\right)\sin\left(\frac{\pi m\ell}{n}\right)=0\\&\sum_{\ell=0}^{2n-1}\sin\left(\frac{\pi k\ell}{n}\right)\sin\left(\frac{\pi m\ell}{n}\right)=\begin{cases}0 & : \ k\neq m\\ n & : \ k=m\end{cases}\end{aligned}}$$

（選点直交性）

が成立する。これを**選点直交性**という。この性質を使うと

$$\sum_{\ell=0}^{2n-1}f_{2n}(x_\ell)\cos\left(\frac{\pi m\ell}{n}\right)=\sum_{\ell=0}^{2n-1}\left[A_0+2\sum_{k=0}^{n-1}\left\{A_k\cos\left(\frac{\pi k\ell}{n}\right)+B_k\sin\left(\frac{\pi k\ell}{n}\right)\right\}+A_n\cos(\pi\ell)\right]\cos\left(\frac{\pi m\ell}{n}\right)=2nA_m$$

より、両辺を $2n$ で割ると

$$\frac{1}{2n}\sum_{\ell=0}^{2n-1}f_{2n}(x_\ell)\cos\left(\frac{\pi m\ell}{n}\right)=A_m$$

が成り立つ。同様にして

$$\frac{1}{2n}\sum_{\ell=0}^{2n-1}f_{2n}(x_\ell)\sin\left(\frac{\pi m\ell}{n}\right)=B_m$$

が成り立つ。そこで

$$\boxed{f_{2n}(x_\ell)=y_\ell} \quad (\ell=0,\ 1,\ \cdots,\ 2n-1)$$

が成立するためには

$$(*)\quad \boxed{A_m=\frac{1}{2n}\sum_{\ell=0}^{2n-1}y_\ell\cos\left(\frac{\pi m\ell}{n}\right)\quad,\quad b_m=\frac{1}{2n}\sum_{\ell=0}^{2n-1}y_\ell\sin\left(\frac{\pi m\ell}{n}\right)}$$

でなければいけない。

逆に $(*)$ であれば選点直交性を用いて $f_{2n}(x_\ell)=y_\ell$ が証明できる。(証明略)

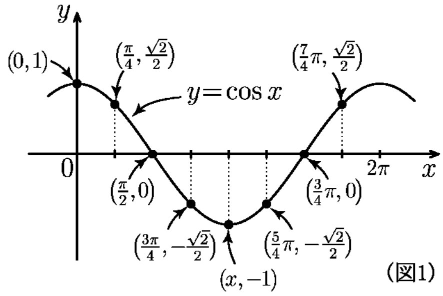
＜三角多項式補間 7＞

このページと次のページでは今までの補間三角多項式 $f_N(x)$ 以外にも補間三角多項式が存在することを示す。ここでは 9 ページの場合 $(\alpha = 0,\ \beta = 2\pi)$ について考える。

例 1 図 1 の 8 点を通る三角多項式 $f_8(x)$ を 9 ページの式で計算すると、

$A_0 = 0,\ A_1 = \frac{1}{2},\ A_2 = A_3 = A_4 = B_1 = B_2 = B_3 = 0$

となり、9 ページ (2) 式にあてはめると

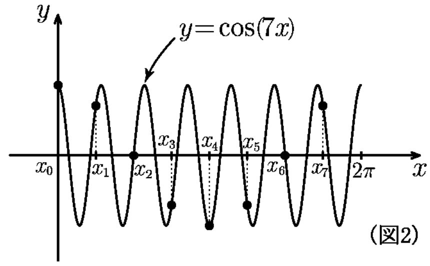


(図1)

$$f_8(x) = A_0 + 2\sum_{k=1}^{3}\left\{A_k \cos(kx) + B_k \sin(kx)\right\} + A_4 \cos(4x) = \cos x$$

となる。

ところがこれ以外にも同じ 8 点を通る三角関数がある。実は $y = \cos(7x)$ のグラフも上と同じ 8 点を通る。(図 2) すなわち



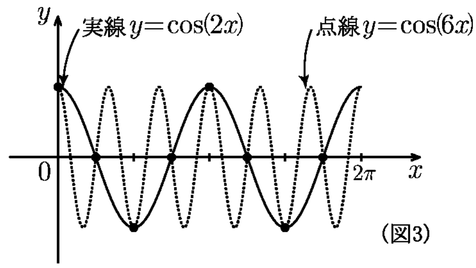
(図2)

$$\boxed{\cos(x_\ell) = \cos(7x_\ell)} \quad (\ell = 0, 1, \cdots, 7)$$

が成り立つ。ただし $x_\ell = \frac{2\pi\ell}{8} = \frac{\pi\ell}{4}$ である。

(注) 9 ページの式からこの関数 $y = \cos(7x)$ はでない。9 ページの補間公式は、できるだけ周期の長い三角多項式による補間公式である。

例 2 例 1 と同じ 8 等分点 $x_\ell(\ell = 0, 1, \cdots, 7)$ に対し $y = \cos(2x)$ と $y = \cos(6x)$ のグラフは同じ点を通る。(図 3) すなわち

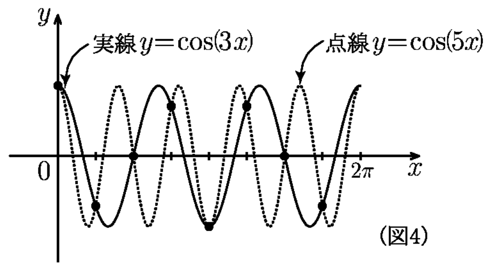


(図3)

$$\boxed{\cos(2x_\ell) = \cos(6x_\ell)} \quad (\ell = 0, 1, \cdots, 7)$$

が成り立つ。

例 3 上と同じ 8 等分点 $x_\ell(\ell = 0, 1, \cdots, 7)$ に対し $y = \cos(3x)$ と $y = \cos(5x)$ のグラフは同じ点を通る。(図 3) すなわち



(図4)

$$\boxed{\cos(3x_\ell) = \cos(5x_\ell)} \quad (\ell = 0, 1, \cdots, 7)$$

が成り立つ。

上の例から 0 から 2π までの 8 等分点 $x_0, x_1, \cdots, x_7$ に対し $\boxed{\cos(kx_\ell) = \cos\big((8-k)x_\ell\big)}$ が成り立つ。一般に 0 から 2π までの N 等分点 $x_0, x_1, \cdots, x_{N-1}$ に対し

$$\boxed{\cos(kx_\ell) = \cos\big((N-k)x_\ell\big)} \quad (\ell = 0, 1, \cdots, N-1)$$

が成り立つ。ただし $x_\ell = \frac{2\pi\ell}{N}$ である。

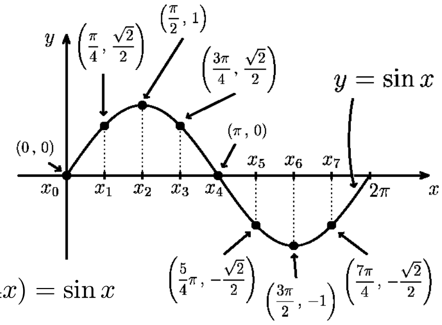
＜ 三角多項式補間 8 ＞

前ページと同じく 9 ページの $f_N(x)$ 以外の補間三角多項式の存在を示す。

例 1　図 1 の 8 点を通る三角多項式 $f_8(x)$ を 9 ページの式で計算すると

$A_0 = A_1 = A_2 = A_3 = A_4 = 0,\ B_1 = \dfrac{1}{2},\ B_2 = B_3 = 0$

となり、9 ページ (2) 式にあてはめると



(図1)

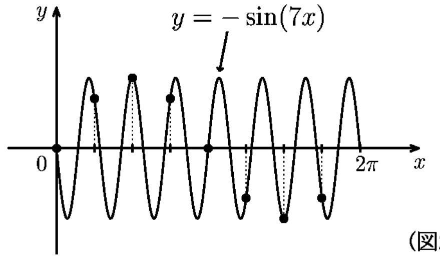
$$f_8(x) = A_0 + 2\sum_{k=1}^{3}\{A_k\cos(kx) + B_k\sin(kx)\} + A_6\cos(4x) = \sin x$$

となる。

ところがこれ以外にも同じ 8 点を通る三角関数がある。実は $y = -\sin(7x)$ のグラフも上と同じ 8 点を通る。(図 2)　すなわち

$$\boxed{\sin(x_\ell) = -\sin(7x_\ell)} \quad (\ell = 0, 1, \cdots, 7)$$

がなりたつ。ただし $x_\ell = \dfrac{2\pi\ell}{8} = \dfrac{\pi\ell}{4}$ である。



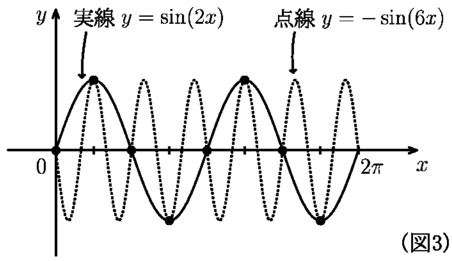
(図2)

(注)　9 ページの式からこの関数 $y = -\sin(7x)$ は導けない。9 ページの補間公式では $N = 8$ の場合一番短い周期が $\big(\cos(4x)$ の周期$\big)\ \dfrac{2\pi}{4} = \dfrac{\pi}{2}$ であり、それより短い周期の三角関数はふくまれない $\big(-\sin(7x)$ の周期は $\dfrac{2\pi}{7}$ である$\big)$。

例 2　例 1 と同じ 8 等分点 $x_\ell(\ell = 0, 1, \cdots, 7)$ に対し $y = \sin(2x)$ と $y = -\sin(6x)$ のグラフは同じ点を通る。(図 3) すなわち

$$\boxed{\sin(2x_\ell) = -\sin(6x_\ell)} \quad (\ell = 0, 1, \cdots, 7)$$

がなりたつ。

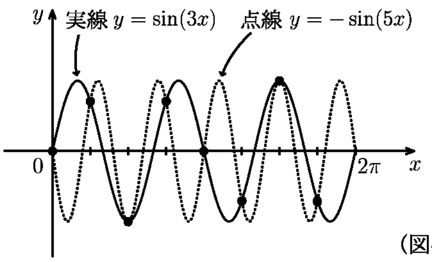


(図3)

例 3　上と同じ 8 等分点 $x_\ell(\ell = 0, 1, \cdots, 7)$ に対し $y = \sin(3x)$ と $y = -\sin(5x)$ のグラフは同じ点を通る。(図 4) すなわち

$$\boxed{\sin(3x_\ell) = -\sin(5x_\ell)} \quad (\ell = 0, 1, \cdots, 7)$$

がなりたつ。



(図4)

上の例から 0 から 2π までの 8 等分点 $x_\ell(\ell = 0, 1, \cdots, 7)$ に対して $\boxed{\sin(kx_\ell) = -\sin\big((8-k)x_\ell\big)}$ がなりたつ。一般に 0 から 2π までの N 等分点 $x_0, x_1, \cdots, x_{N-1}$ に対して

$$\boxed{\sin(kx_\ell) = -\sin\big((N-k)x_\ell\big)} \quad (\ell = 0, 1, \cdots, N-1)$$

がなりたつ。ただし $x_\ell = \dfrac{2\pi\ell}{N}$ である。

< 三角多項式補間 9 >

自然数 N , k , ℓ に対し次の式が成立する。

(1) $\boxed{\cos\left(\dfrac{2\pi k\ell}{N}\right)=\cos\left(\dfrac{2\pi(N-k)\ell}{N}\right)}$, (2) $\boxed{\sin\left(\dfrac{2\pi k\ell}{N}\right)=-\sin\left(\dfrac{2\pi(N-k)\ell}{N}\right)}$

これらの式は 15 , 16 ページの下の式で $x_\ell=\dfrac{2\pi\ell}{N}$ を代入した式である。

(証明)

(1) $\cos\left(\dfrac{2\pi(N-k)\ell}{N}\right)=\cos\left(2\pi\ell-\dfrac{2\pi k\ell}{N}\right)=\cos\left(-\dfrac{2\pi k\ell}{N}\right)=\cos\left(\dfrac{2\pi k\ell}{N}\right)$

(2) $-\sin\left(\dfrac{2\pi(N-k)\ell}{N}\right)=-\sin\left(2\pi\ell-\dfrac{2\pi k\ell}{N}\right)=-\sin\left(-\dfrac{2\pi k\ell}{N}\right)=\sin\left(\dfrac{2\pi k\ell}{N}\right)$

13 ページの A_k , B_k の定義式

$$(*)\quad \boxed{A_k=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\cos\left(\frac{2\pi k\ell}{N}\right)\quad,\quad B_k=\frac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\sin\left(\frac{2\pi k\ell}{N}\right)}$$

と上の (1) , (2) 式から以下の等式が成立する。

(3) $\boxed{A_k=A_{N-k}}$, (4) $\boxed{B_k=-B_{N-k}}$, (5) $\boxed{B_0=0}$

(証明)

(3) $A_k=\dfrac{1}{N}\displaystyle\sum_{\ell=0}^{N-1}y_\ell\cos\left(\dfrac{2\pi k\ell}{N}\right)=\dfrac{1}{N}\sum_{\ell=0}^{N-1}y_\ell\cos\left(\dfrac{2\pi(N-k)\ell}{N}\right)=A_{N-k}$

(5) $B_0=\dfrac{1}{N}\displaystyle\sum_{\ell=0}^{N-1}y_\ell\sin(0)=0$

問 1 (4) 式を証明せよ。

問 2 $N=2n$ (偶数) の場合は $(*)$ 式は

$$A_k=\frac{1}{2n}\sum_{\ell=0}^{2n-1}y_\ell\cos\left(\frac{\pi k\ell}{n}\right)\quad,\quad B_k=\frac{1}{2n}\sum_{\ell=0}^{2n-1}y_\ell\sin\left(\frac{\pi k\ell}{n}\right)$$

となる。このとき

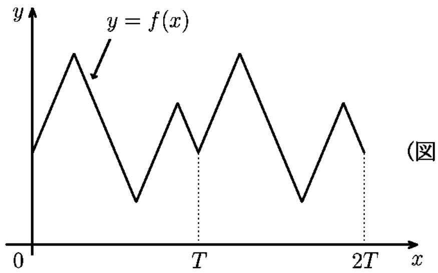
$$(6)\quad \boxed{B_n=0}$$

を証明せよ。

＜ フーリェ近似 1 ＞

周期関数を三角多項式で近似することを**フーリェ近似**という。

例　$f(x)$ が図 1 のような周期関数であるとき、
$f(x)$ を 13 ページの三角多項式 $f_N(x)$ で近似したい。

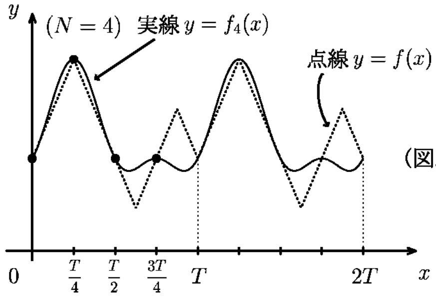


(図1)

(1)　図 2 のように 0 から T までを 4 等分し、$y = f(x)$ 上の 4 点

$$\Big(0, f(0)\Big), \left(\frac{T}{2}, f\left(\frac{T}{4}\right)\right), \left(\frac{T}{2}, f\left(\frac{T}{2}\right)\right), \left(\frac{3T}{4}, f\left(\frac{3T}{4}\right)\right)$$

を通る三角多項式が $y = f_4(x)$(図 2 の実線) である。

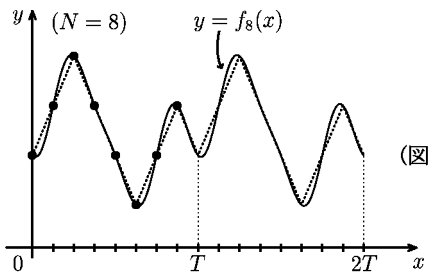


(図2)

(2)　図 3 のように 0 から T までを 8 等分し、$y = f(x)$ 上の 8 点

$$\left(\frac{T\ell}{8}, f\left(\frac{T\ell}{8}\right)\right) \quad (\ell = 0, 1, \cdots, 7)$$

を通る三角多項式が $y = f_8(x)$(図 3 の実線) である。



(図3)

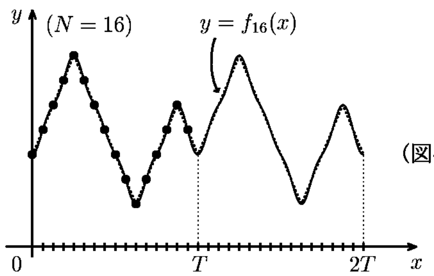
(3)　図 4 のように 0 から T までを 16 等分し、$y = f(x)$ 上の 16 点

$$\left(\frac{T\ell}{16}, f\left(\frac{T\ell}{16}\right)\right) \quad (\ell = 0, 1, \cdots, 15)$$

を通る三角多項式が $y = f_{16}(x)$(図 4 の実線) である。
元の周期関数 $y = f(x)$(図 4 の点線) は
三角多項式 $y = f_{16}(x)$ とほとんど一致している。
従ってこの場合は $f(x)$ を $f_{16}(x)$ で近似できる。
すなわち

$$f(x) \fallingdotseq f_{16}(x)$$

と考えてよい。



(図4)

一般に周期 T の周期関数 $y = f(x)$ に対し、この曲線上の N 点

$$\left(\frac{T\ell}{N}, f\left(\frac{T\ell}{N}\right)\right) \quad (\ell = 0, 1, \cdots, N-1)$$

を通る三角多項式 $f_N(x)$ を考える。N が十分大きい数ならば、もとの周期関数 $f(x)$ を $f_N(x)$ で近似できる。すなわち

$$\boxed{f(x) \fallingdotseq f_N(x)} \quad \text{(フーリェ近似)}$$

と考えてよい。ここで $f_N(x)$ は 13 ページで

$$\alpha = 0\ ,\ \beta = T\ ,\ y_\ell = f\left(\frac{T\ell}{N}\right) \quad (\ell = 0, 1, \cdots, N-1)$$

の場合の三角多項式 $f_N(x)$ である。

＜ フーリエ近似2 ＞

フーリェ近似式 $f(x) \fallingdotseq f_N(x)$ は N が大きくないと成立しない。たとへば前ページの例で $N=4$ の場合、$f_4(x)$ が $f(x)$ を近似しているとは言えない。N がどの程度大きい数でなければならないかを例をあげて説明する。

例 1

$T=2\pi,\quad f(x)=\sin(5x)$

の場合の $f_N(x)$ は以下のようになる。

$N=6$ のとき $f_6(x)=-\sin x$ (図 2)

$N=8$ のとき $f_8(x)=-\sin(3x)$ (図 3)

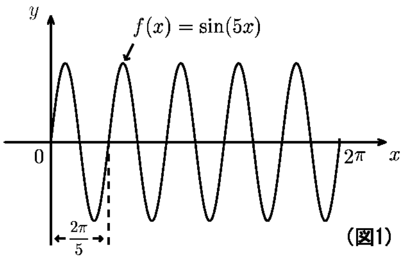
$N=10$ のとき $f_{10}(x)=0$ (図 4)

$N=11$ のとき $f_{11}(x)=\sin(5x)$ (図 5)

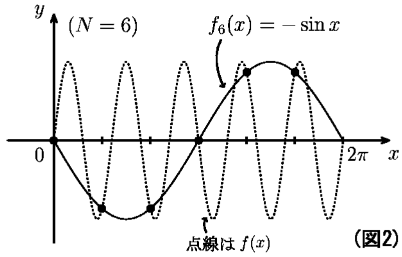
$N=12$ のとき $f_{12}(x)=\sin(5x)$ (図 6)

これより N が 10 以下のときは $f_N(x)$ は $f(x)$ を近似しない。

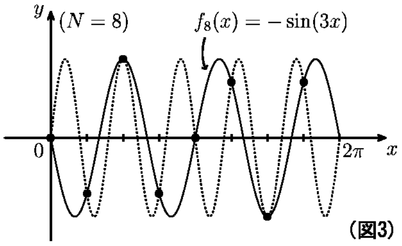
$N \geqq 11$ のとき $f_N(x)=f(x)$ となる。



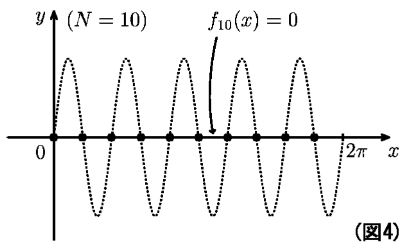
(図1)



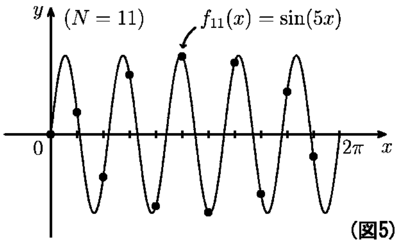
(図2)



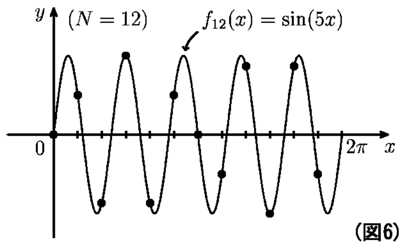
(図3)



(図4)



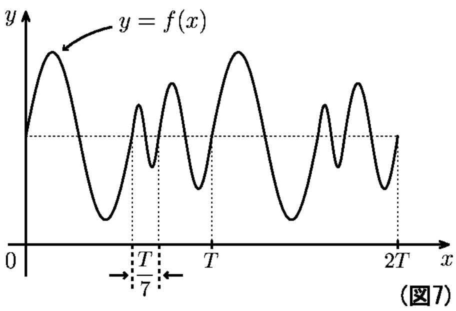
(図5)



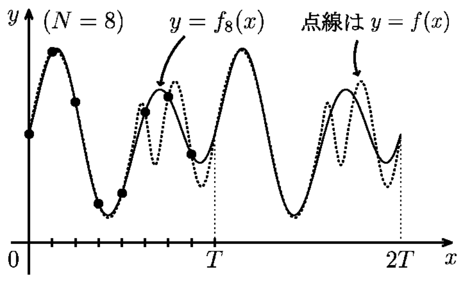
(図6)

例 2

$f(x)$ は図 7 のような周期関数とする。曲線 $y=f(x)$ は 0 から T までの間に 3 つの波があり、まん中の小さな波の周期は $\frac{T}{7}$ である。$f(x)$ を $f_N(x)$ で近似したい。



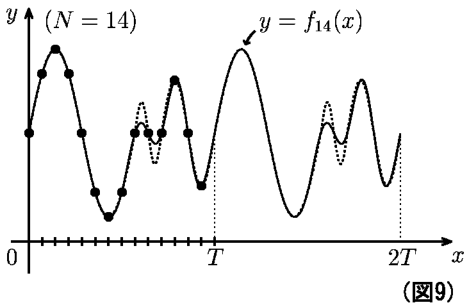
(図7)



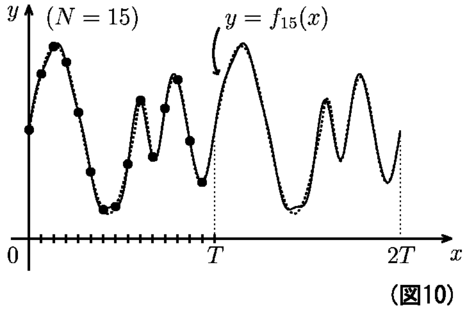
(図8)

N が 14 以下の場合は $f_N(x)$ はまん中の小さな波を近似しない (図 8, 図 9)。

$f(x)$ を $f_N(x)$ で近似するためには、N が 15 以上でなければならない (図 10)。



(図9)



(図10)

＜ 離散フーリエ変換 1 ＞

$N = 2n$(偶数) のときフーリエ近似式 $f_{2n}(x)$ の $x = x_\ell$ での値は 14 ページより

$$(1) \qquad \boxed{y_\ell = f_{2n}(x_\ell) = A_0 + 2\sum_{k=1}^{n-1}\left\{A_k \cos\left(\tfrac{\pi k\ell}{n}\right) + B_k \sin\left(\tfrac{\pi k\ell}{n}\right)\right\} + A_n \cos(\pi\ell)}$$

であった。ここで A_k, B_k の定義式

$$A_k = \frac{1}{2n}\sum_{\ell=0}^{2n-1} y_\ell \cos\left(\tfrac{\pi k\ell}{n}\right), \quad B_k = \frac{1}{2n}\sum_{\ell=0}^{2n-1} y_\ell \sin\left(\tfrac{\pi k\ell}{n}\right)$$

から以下の等式が成立する (17 ページ参照)。

$$(2) \qquad \boxed{\begin{array}{c} A_{2n-k} = A_k \quad , \quad B_{2n-k} = -B_k \qquad (k = 1, 2, \cdots, 2n-1) \\[2ex] B_0 = 0 \quad , \quad B_n = 0 \end{array}}$$

ここで、オイラーの公式

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta} \qquad (i = \sqrt{-1}\ \text{は虚数単位})$$

から、三角関数の複素指数表示

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \quad , \quad \sin\theta = \left(\frac{e^{-i\theta} - e^{i\theta}}{2}\right)i$$

を用いると、(1) 式は以下のように書きなおせる。

$$(1)' \qquad \boxed{y_\ell = \sum_{k=1}^{2n-1} (A_k - B_k i)e^{\frac{\pi k\ell}{n}i}}$$

$(1)'$ の右辺は複素数を使っているが、((2) 式を使って) 計算すると実数になり (1) 式と同じ式になる。

問 (2) 式を使うことによって、(1) 式から $(1)'$ 式を導け。

＜離散フーリエ変換2＞

$N=2n$ (偶数) のときフーリエ近似式 $f_N(x)$ の $x=x_\ell$ での値は前ページより

(1) $$\boxed{y_\ell = f_N(x_\ell) = \sum_{k=0}^{N-1}(A_k - B_k i)e^{\frac{2\pi k\ell}{N}i}}$$

と書ける。ただし A_k , B_k は

(2) $$\boxed{A_k = \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell \cos\left(\frac{2\pi k\ell}{N}\right),\ B_k = \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell \sin\left(\frac{2\pi k\ell}{N}\right)}$$

となる。

(注) $N=2n-1$(奇数) のときも全く同じ形に書きなおせる。

ここで

$$\boxed{C_k = A_k - B_k i}\ (k=0,1,\cdots,N-1)$$

とおくと (2) 式より

$$\begin{aligned} C_k &= \left(\frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell \cos\left(\frac{2\pi k\ell}{N}\right)\right) - \left(\frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell \sin\left(\frac{2\pi k\ell}{N}\right)\right) i \\ &= \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell \left\{\cos\left(\frac{2\pi k\ell}{N}\right) - i\sin\left(\frac{2\pi k\ell}{N}\right)\right\} \\ &= \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell e^{-\frac{2\pi k\ell}{N}i} \end{aligned}$$

と書ける。この式

$$\boxed{C_k = \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell e^{-\frac{2\pi k\ell}{N}i}}\ \text{(離散フーリエ変換)}$$

を**離散フーリエ変換** (Discrete Fourier Tranform) という。
また (1) 式は

$$\boxed{y_\ell = \sum_{k=0}^{N-1} C_k e^{\frac{2\pi k\ell}{N}i}}\ \text{(離散フーリエ逆変換)}$$

と書ける。この式を**離散フーリエ逆変換** (Inverse Discrete Frourier Transform) という。

< 離散フーリエ変換3 >

N 個の実数 $y_0,\ y_1,\ \cdots,\ y_{N-1}$ と N 個の複素数 $C_0,\ C_1,\ \cdots,\ C_{N-1}$ に対して以下の等式が成立する。(証明略)

$$(*)\qquad \boxed{\sum_{\ell=0}^{N-1}\left|y_\ell - \sum_{k=0}^{N-1} C_k e^{\frac{2\pi k\ell}{N}i}\right|^2 = N\sum_{k=0}^{N-1}\left|C_k - \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell e^{-\frac{2\pi k\ell}{N}i}\right|^2}$$

この等式より以下のことがわかる。

もし全ての k $(k=0,\ 1,\ \cdots,\ N-1)$ に対して

$$(**)\qquad \boxed{C_k = \frac{1}{N}\sum_{\ell=0}^{N-1} y_\ell e^{-\frac{2\pi k\ell}{N}i}}\quad \text{(離散フーリエ変換)}$$

が成り立てば

$\Downarrow$

$(*)$ 式の右辺=0

$\Downarrow$

$(*)$ 式の左辺=0

$\Downarrow$

$$(***)\qquad \boxed{y_\ell = \sum_{k=0}^{N-1} C_k e^{\frac{2\pi k\ell}{N}i}}\quad \text{(離散フーリエ逆変換)}$$

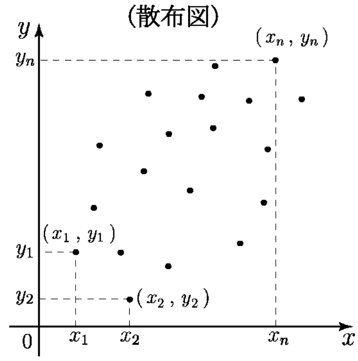
が全ての ℓ $(\ell=0,\ 1,\ \cdots,\ N-1)$ に対して成立する。

逆に $(***)$ 式が成立すれば、$(**)$ 式も成立する。

(注) $(*)$ 式の左辺は y_ℓ と $\displaystyle\sum_{k=0}^{N-1} C_k e^{\frac{2\pi k\ell}{N}i}$ との距離の 2 乗の和を意味している。2 乗の和を最小にするような方法を**最小 2 乗法**という。

問　$(**)$ 式から $C_{N-k} = \overline{C_k}$(C_k の複素共役) を導け。

＜ 2 変数データの統計量 ＞



2 変数のデータ (x_1, y_1) , (x_2, y_2) , $\cdots$, (x_n, y_n) を右図のように座標平面上の点として表したものを**散布図**という。

x 成分に関する平均と分散を

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) \quad : \quad x \text{ の平均}$$

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\Big\{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2\Big\} \quad : \quad x \text{ の分散}$$

と書く。同様に y 成分に関する平均と分散を

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}(y_1 + y_2 + \cdots + y_n) \quad : \quad y \text{ の平均}$$

$$S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\Big\{(y_1-\bar{y})^2 + (y_2-\bar{y})^2 + \cdots + (y_n-\bar{y})^2\Big\} \quad : \quad y \text{ の分散}$$

と書く。次に x 成分と y 成分の表す関係を意味する量として

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{n}\Big\{(x_1-\bar{x})(y_1-\bar{y}) + \cdots + (x_n-\bar{x})(y_n-\bar{y})\Big\} \quad : x \text{ と } y \text{ の共分散}$$

を x と y の**共分散** (covariance) といい、$S_{xy} = Cov(x, y)$ とも書くことがある。また

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \frac{\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \times \left(\displaystyle\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}} \quad : \quad x \text{ と } y \text{ の相関係数}$$

を x と y の**相関係数** (correlation coefficient) といい、$r_{xy} = Corr(x, y)$ とも表すことがある。

例　5 人の身長 (x) と体重 (y) が右の表の場合、

身長 x (cm)	体重 y (kg)
150	45
160	50
165	55
170	60
180	70

$$\bar{x} = \frac{1}{5}(150 + 160 + 165 + 170 + 180) = 165 \quad : \quad \text{身長の平均}$$

$$\bar{y} = \frac{1}{5}(45 + 50 + 55 + 60 + 70) = 56 \quad : \quad \text{体重の平均}$$

$$S_{xx} = \frac{1}{5}\Big\{(150-165)^2+(160-165)^2+(165-165)^2+(170-165)^2+(180-165)^2\Big\} = 100 \quad (x \text{ の分散})$$

$$S_{yy} = \frac{1}{5}\Big\{(45-56)^2 + (50-56)^2 + (55-56)^2 + (60-56)^2 + (70-56)^2\Big\} = 74 \quad (y \text{ の分散})$$

$$S_{xy} = \frac{1}{5}\Big\{(150-165)(45-56) + (160-165)(50-56) + (165-165)(55-56) + (170-165)(60-56) + (180-165)(70-56)\Big\} = 85 \quad (\text{共分数})$$

より相関係数は

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \frac{85}{\sqrt{100 \times 74}} = \frac{17}{2\sqrt{94}} \fallingdotseq 0.988 \quad (\text{相関係数})$$

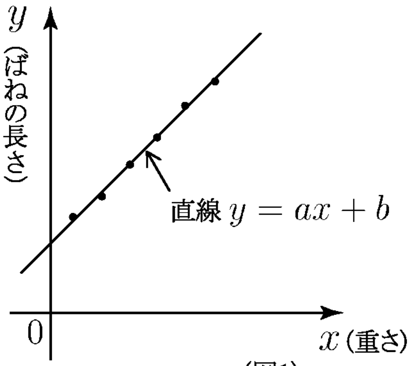
＜回帰直線 1 ＞

2 変数データ $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ の散布図から x と y の関係がはっきりわかる場合とそうでない場合がある。

例 1 ばねに重りをつるしたとき、ばねの重さを x 成分、ばねの長さを y 成分として計ったデータの散布図は図 1 のようにほぼ直線的に並ぶ。この直線の傾きを a, y 切片を b とすれば

$$y = ax + b$$

の関係がわかる。ばね計りを作るときはこの関係式を求めてから、ばね計りの「めもり」を決めればよい。



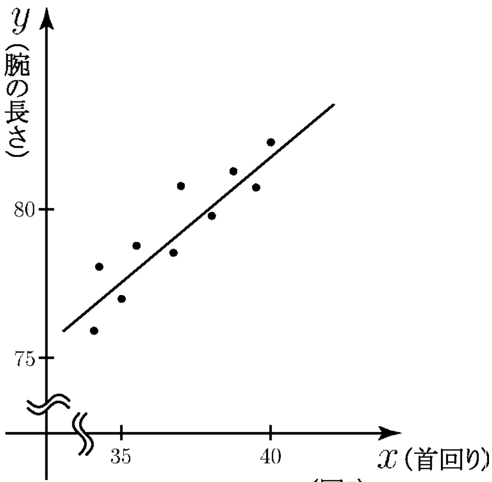
(図1)

例 2 何人かの「首周りの長さ」と「腕の長さ」を計った。「首周りの長さ」を x 成分、「腕の長さ」を y 成分として計ったデータの散布図は図 2 のようになった。図 2 を見ると首周りが大きい程腕の長さが長い傾向が ある。しかし x と y の関係はわからない。

もし平均的な体形の人の x と y の関係がわかれば便利である。そうすればシャツメーカーが全てのサイズのシャツを作らなくてもすむ。たとえば平均的な体形の人の x と y の関係が

$$y = f(x)$$

とすれば、首周りが x (cm) の場合、腕の長さは $f(x) \pm 3\,\mathrm{cm}$ 程度のサイズをそろえておけばよいであろう。



(図2)

このような関数 $f(x)$ を求め、その関数と元のデータとの関係を調べる分野を**「回帰分析」**という。

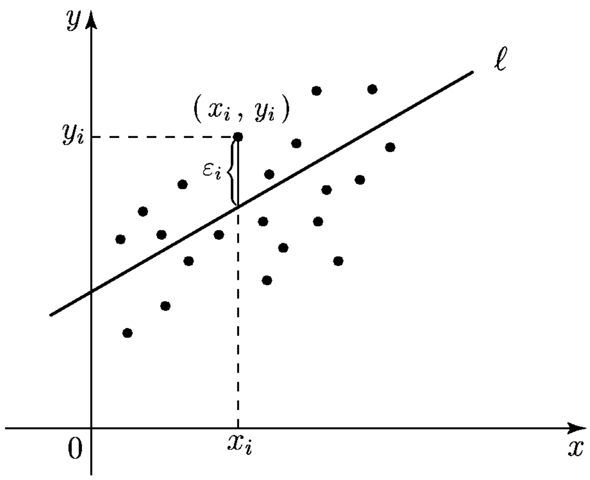
特に $f(x)$ が x の一次式の場合、$y = f(x)$ のグラフは図 2 のような直線を表す。この「直線をどのように定めるべきか？」という問題は元のデータ x と y の関係をどのように判断するかによって決まる。

この本では「回帰直線」(次ページ) と「直交回帰直線」(37 ページ) を解説する。

＜回帰直線 2＞

n 個の 2 変数データ
$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$
の散布図が右図のような場合、第 i 番目の点 (x_i, y_i) と直線 ℓ との y 軸方向の距離を ε_i とする。
この距離の 2 乗の和

$$\sum_{i=1}^{n} \varepsilon_i^2$$



を最小にするような直線 ℓ をこのデータの**回帰直線**という。
この直線 ℓ の方程式は元のデータ $(x_1, y_1), \cdots, (x_n, y_n)$ から以下のように求まる。

直線 ℓ の方程式 ： $\boxed{y = \dfrac{S_{xy}}{S_{xx}}(x - \bar{x}) + \bar{y}}$ (回帰直線の方程式)

ただし

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (x \text{ の平均}) \quad , \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (y \text{ の平均})$$

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (x \text{ の分散}) \quad , \quad S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad (\text{共分散})$$

である。この方程式を導くためには距離の 2 乗の和 $\sum_{i=1}^{n}\varepsilon_i^2$ を最小にするような直線を求める (27 ページに証明する)。この方法を**最小 2 乗法**という。
この回帰直線は y 成分が x 成分の関数 (一次式) によって表されるという意味で**「y の x に対する回帰直線」**とも言う。この直線は点 $(\bar{x}, \bar{y})$ を通り、傾き S_{xy}/S_{xx} の直線である。

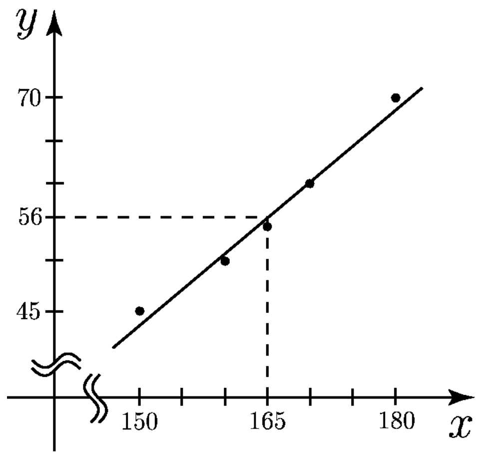
例　23 ページの例の場合は

$$\bar{x} = 165\ ,\ \bar{y} = 56\ ,\ S_{xx} = 100\ ,\ S_{xy} = 85$$

より回帰直線は

$$y = \frac{85}{100}(x - 165) + 56$$

であるから点 $(\bar{x}, \bar{y}) = (165, 56)$ を通り 傾き 0.85 の直線である。(右図参照)。



< 回帰直線 3 >

例 10 人の「首回りの長さ」と「腕の長さ」を計った。x 成分を首回り、y 成分を腕の長さとして 10 人のデータが以下のようになった。

$(38, 81)$, $(40, 82)$, $(34, 78)$, $(41, 81)$, $(34, 75)$
$(38, 79)$, $(42, 83)$, $(36, 79)$, $(35, 77)$, $(39, 80)$

首回りの平均は

$$\bar{x} = \frac{1}{10}\Big\{38 + 40 + 34 + 41 + 34 + 38 + 42 + 36 + 35 + 39\Big\} = 37.7 \quad \text{(首回りの平均)}$$

腕の長さの平均は

$$\bar{y} = \frac{1}{10}\Big\{81 + 82 + 78 + 81 + 75 + 79 + 83 + 79 + 77 + 80\Big\} = 79.5 \quad \text{(腕の長さの平均)}$$

x 成分の分散は

$$S_{xx} = \frac{1}{10}\Big\{(38 - 37.7)^2 + (40 - 37.7)^2 + \cdots + (39 - 37.7)^2\Big\} = 7.41 \quad (x \text{ の分散})$$

y 成分の分散は

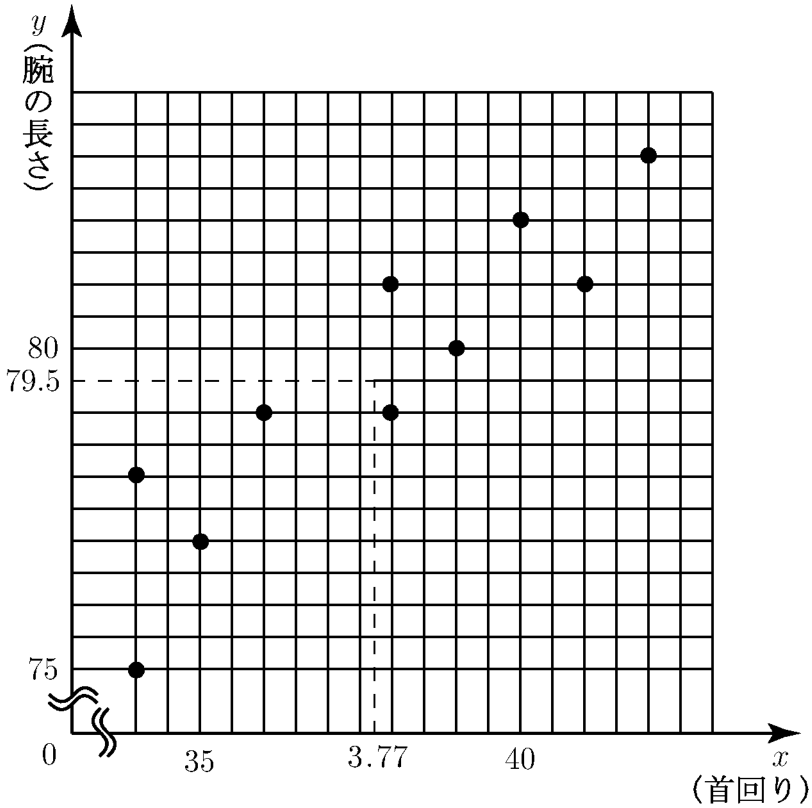
$$S_{yy} = \frac{1}{10}\Big\{(81 - 79.5)^2 + (82 - 79.5)^2 + \cdots + (80 - 79.5)^2\Big\} = 5.25 \quad (y \text{ の分散})$$

共分散は

$$S_{xy} = \frac{1}{10}\Big\{(38 - 37.7)(81 - 79.5) + \cdots + (39 - 37.7)(80 - 79.5)\Big\} = 5.65 \quad \text{(共分散)}$$

となる。

問 上の例の場合に回帰直線の方程式を求めよ。
また右の散布図上に回帰直線のグラフを描け。



＜回帰直線の導出＞

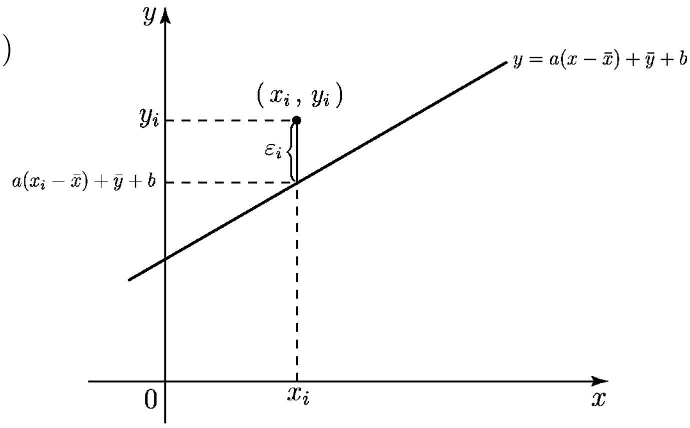
n 個の 2 変数データ

$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$

に対する回帰直線の方程式が

$$\boxed{y = \frac{S_{xy}}{S_{xx}}(x - \bar{x}) + \bar{y}}$$

であることを証明する。



＜ **証明** ＞

求める回帰直線を仮に

$$y = a(x - \bar{x}) + \bar{y} + b$$

とおいて、 $\boxed{a = \frac{S_{xy}}{S_{xx}}}$, $\boxed{b = 0}$ であることを示す。

第 i 番目の点 (x_i, y_i) から直線への y 軸方向の距離 ε_i は図より

$$\varepsilon_i = y_i - \left\{a(x_i - \bar{x}) + \bar{y} + b\right\} = y_i - \bar{y} - a(x_i - \bar{x}) - b$$

であるから

$$\begin{aligned}\sum_{i=1}^{n} \varepsilon_i^2 &= \sum_{i=1}^{n} \left\{(y - \bar{y}) - a(x_i - \bar{x}) - b\right\}^2 \\ &= \sum_{i=1}^{n} \left\{(y_i - \bar{y})^2 + a^2(x_i - \bar{x})^2 + b^2 - 2a(y_i - \bar{y})(x_i - \bar{x}) - 2b(y_i - \bar{y}) + 2ab(x_i - \bar{x})\right\}\end{aligned}$$

となる。ここで

$$\sum_{i=1}^{n} 2ab(x_i - \bar{x}) = 2ab\left\{\sum_{i=1}^{n}(x_i - \bar{x})\right\} = 2ab\left\{\left(\sum_{i=1}^{n} x_i\right) - n\bar{x}\right\} = 2ab\left\{\left(\sum_{i=1}^{n} x_i\right) - n \times \frac{1}{n}\sum_{i=1}^{n} x_i\right\} = 0$$

である。同様にして $\displaystyle\sum_{i=1}^{n} 2b(y_i - \bar{y}) = 0$ となるので $\displaystyle\sum_{i=1}^{n} \varepsilon_i^2$ は

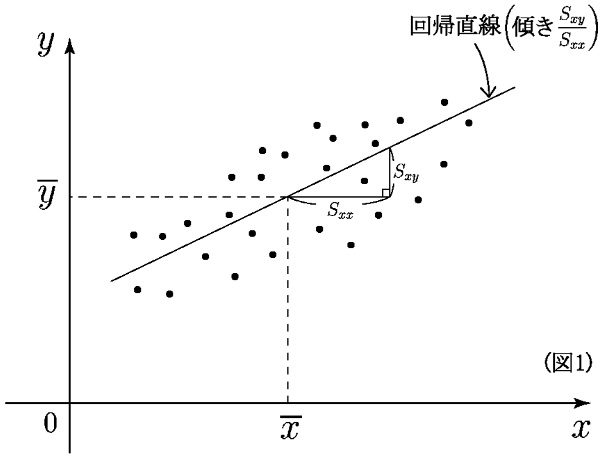
$$\begin{aligned}\sum_{i=1}^{n} \varepsilon_i^2 &= \left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right) a^2 - 2\left(\sum_{i-1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right) a + \left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right) + nb^2 \\ &= \left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)\left\{a - \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2}\right\}^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{\left(\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2} + nb^2\end{aligned}$$

という a の 2 次式の形になる。従って $\displaystyle\sum_{i=1}^{n} \varepsilon_i^2$ は

$$a = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\displaystyle\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} \quad , \qquad b = 0$$

のとき最小になる。(証明終)

＜相関係数＞



(図1)

体重の大きな人は身長も高い傾向にある。このように2つの変数の間に関連があることを**相関** (correlation) があるという。

図1、図2、図4のように散布図が右上がりに並び、2つの変数の大小がそろう傾向にあるとき、2つの変数の間には**正の相関**があるという。逆に図3、図5のように散布図が右下がりに並ぶ傾向にあるときは**負の相関**があるという。また図6のように右上がりの傾向も右下がりの傾向も見えないときは**無相関**であるという。

この相関の強さを表す値が**相関係数**である。n 個の2変数データ (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) の相関係数は

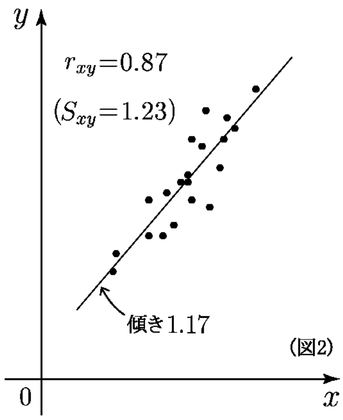
$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(\displaystyle\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}}$$

(相関係数)

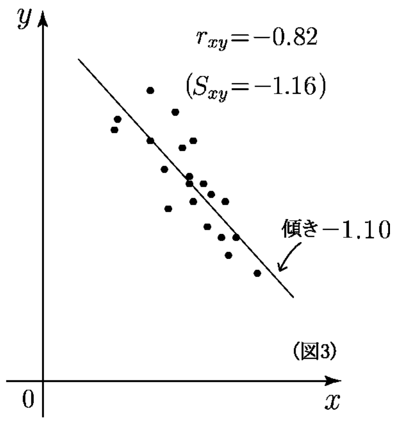
である。相関係数の符号は共分散 $S_{xy} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$ のプラス・マイナスによって決まる。また回帰直線の傾き $\dfrac{S_{xy}}{S_{xx}}$ の符号も ($S_{xx} > 0$ より) 共分散 S_{xy} のプラス・マイナスによって定まる。次ページで証明するが相関係数は常に

$$-1 \leqq r_{xy} \leqq 1$$

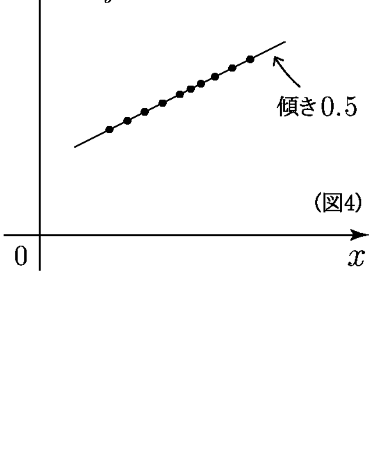
である。



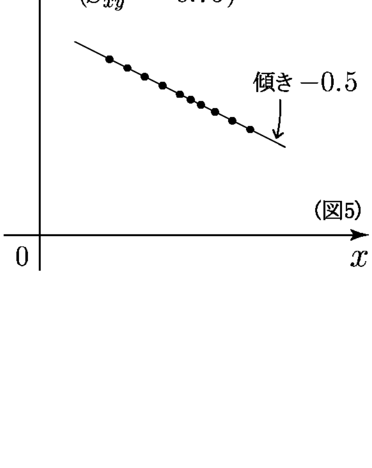
(図2)



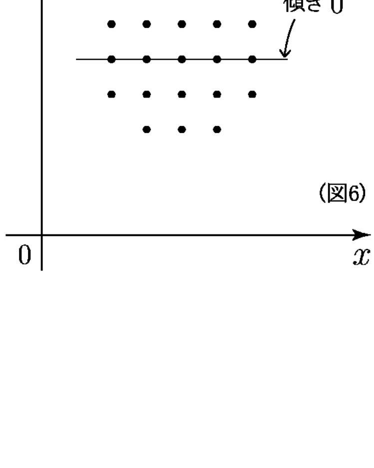
(図3)



(図4)

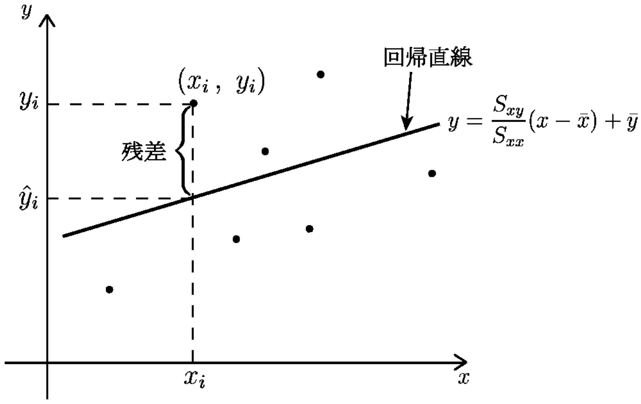


(図5)



(図6)

< 残差平方和 >



n 個の 2 変数データ

$(x_1\,,\ y_1)\,,\ (x_2\,,\ y_2)\,,\ \cdots,\ (x_n\,,\ y_n)$

の回帰直線を

$$y=\frac{S_{xy}}{S_{xx}}(x-\bar{x})+\bar{y}$$

とする。第 i 番目の点 $(x_i\,,\ y_i)$ に対し、
$x=x_i$ における回帰直線上の点を $(x_i\,,\ \hat{y}_i)$
とすれば $\hat{y}_i=\dfrac{S_{xy}}{S_{xx}}(x_i-\bar{x})+\bar{y}$ であり、 y_i との差 $y_i-\hat{y}_i$ を**残差**という。
この残差の平方の和

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \qquad : \qquad \text{残差平方和}$$

を**残差平方和**という。27 ページの証明より

$$\sum_{i=1}^{n}\left\{y_i-\left(a(x_i-\bar{x})+\bar{y}+b\right)\right\}^2=\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(a-\frac{S_{xy}}{S_{xx}}\right)^2+\sum_{i=1}^{n}(y_i-\bar{y})^2-\frac{\left(\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)^2}{\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2}+nb^2$$

となり、 $a=\dfrac{S_{xy}}{S_{xx}}$, $b=0$ のとき $a(x_i-\bar{x})+\bar{y}+b=\hat{y}_i$ だから

$$\begin{aligned}\sum_{i=1}^{n}\{y_i-\hat{y}_i\}^2&=\sum_{i=1}^{n}(y_i-\bar{y})^2-\frac{\left(\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)^2}{\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2}\\&=\sum_{i=1}^{n}(y_i-\bar{y})^2\left\{1-\frac{\left(\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)^2}{\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}\right\}\end{aligned}$$

となる。ここで相関関数 r_{xy} の定義より

$$(r_{xy})^2=\frac{(S_{xy})^2}{S_{xx}S_{yy}}=\frac{\left(\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)^2}{\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

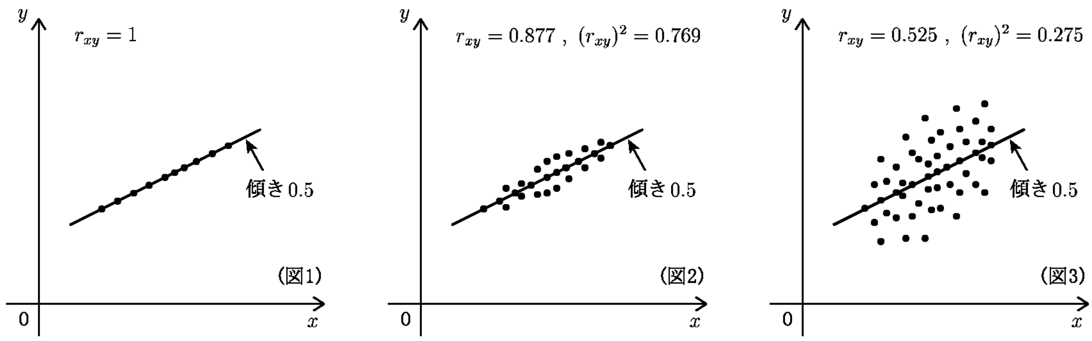
となるから残差平方和は

$$\boxed{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\bar{y})^2\left\{1-(r_{xy})^2\right\}} \qquad \text{(残差平方和)}$$

と書ける。この式から両辺は 0 以上だから $1-{r_{xy}}^2\geqq 0 \ \Rightarrow\ \boxed{-1\leqq r_{xy}\leqq 1}$ がわかる。
また ${r_{xy}}^2=1$ のときは両辺は 0 になるから、このとき散布図上の点は全て回帰直線上に並んでいることがわかる。

＜決定係数＞

例 下の3通りのデータ (散布図) はいずれも回帰直線が同じ (傾き 0.5) である。図1は全ての点が回帰直線上に並んでいる (相関係数 $r_{xy}=1$)。図2はほとんどの点が回帰直線の近くにあり、相関が強い ($r_{xy}=0.877$)。図3は相関が弱い。この場合は回帰直線がこのデータを代表しているとはいえない。しかし相関係数は $r_{xy}=0.525$ でさほど小さくない。
従って相関関数の値だけでは「回帰直線がデータを代表しているかどうか」を判断できない。



回帰直線のあてはまりの尺度を決めたい。
1つの考え方は前ページの残差平方和

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\bar{y})^2\left\{1-(r_{xy})^2\right\} \qquad : \qquad (\text{残差平方和})$$

を計り、この値が小さい程あてはまりが良いと考えられる。しかしこの値は単位のとり方によって変わる。たとえば x 成分が体重 (kg) で、y 成分が身長 (cm) であるとする。もし y 成分の単位を mm に変えると残差は10倍、残差平方和は100倍になる。単位のとり方によらない尺度にするためには残差平方和を $\displaystyle\sum_{i=1}^{n}(y_i-\bar{y})^2$ で割ればよい。実際には

$$\boxed{(r_{xy})^2=1-\frac{\displaystyle\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\displaystyle\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (\text{決定係数})$$

を**決定係数**といい、$(r_{xy})^2$ を回帰式のあてはまりの尺度とする。

< 最小 2 乗法の数理 1 >

多項式回帰や重回帰分析等で用いられる最小 2 乗法は、次の定理の応用である。

< **定理** >

n 個の $m+2$ 変数データ

$$(x_{01}, x_{01}, \cdots, x_{m1}, y_1), \quad (x_{02}, x_{12}, \cdots, x_{m2}, y_2), \quad \cdots, \quad (x_{0n}, x_{1n}, \cdots, x_{mn}, y_n)$$

に対し、連立方程式

$$(*) \begin{cases} \displaystyle\sum_{i=1}^{n} \{y_i - (a_0x_{0i} + a_1x_{1i} + \cdots + a_mx_{mi})\}\, x_{0i} = 0 \\ \displaystyle\sum_{i=1}^{n} \{y_i - (a_0x_{0i} + a_1x_{1i} + \cdots + a_mx_{mi})\}\, x_{1i} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \displaystyle\sum_{i=1}^{n} \{y_i - (a_0x_{0i} + a_1x_{1i} + \cdots + a_mx_{mi})\}\, x_{mi} = 0 \end{cases}$$

の解があるとする。その解を $(a_0, a_1, \cdots, a_m)$ とする。このとき任意の実数 $A_0, A_1, \cdots, A_m$ に対し、次の不等式が成立する。

$$\sum_{i=1}^{n} \{y_i - (A_0x_{0i} + A_1x_{1i} + \cdots + A_mx_{mi})\}^2 \geqq \sum_{i=1}^{n} \{y_i - (a_0x_{0i} + a_1x_{1i} + \cdots + a_mx_{mi})\}^2$$

この連立方程式 $(*)$ を通常**正規方程式**という。$(*)$ を $a_0, a_1, \cdots, a_m$ の連立一次方程式の形に書きなおすと以下のようになる。

$$(*)' \begin{cases} \left(\displaystyle\sum_{i=1}^{n} x_{0i}x_{0i}\right) a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i}x_{0i}\right) a_1 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi}x_{0i}\right) a_m = \displaystyle\sum_{i=1}^{n} y_ix_{0i} \\ \left(\displaystyle\sum_{i=1}^{n} x_{0i}x_{1i}\right) a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i}x_{1i}\right) a_1 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi}x_{1i}\right) a_m = \displaystyle\sum_{i=1}^{n} y_ix_{1i} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \left(\displaystyle\sum_{i=1}^{n} x_{0i}x_{mi}\right) a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i}x_{mi}\right) a_1 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi}x_{mi}\right) a_m = \displaystyle\sum_{i=1}^{n} y_ix_{mi} \end{cases}$$

$(*)$ と $(*)'$ は同じ式であり、両方とも正規方程式という。
この定理により最小 2 乗法は実質的に連立一次方程式 $(*)'$ に帰着する。
連立一次方程式は解がない場合もあるが、本書では解がある場合だけをとりあつかう。連立一次方程式の詳しい理論は線形代数等の本を見てほしい。
具体的に連立一次方程式を解く場合は「数値計算」や「最小 2 乗法」についての詳しい解説本を見てほしい。

＜最小2乗法の数理2＞

前ページの定理を証明する。

[**証明**] $a_0, a_1, \cdots, a_m$ を正規方程式 $(*)$ の解とする。つまり

$$(*)\begin{cases} \displaystyle\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\,x_{0i}=0 \\ \displaystyle\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\,x_{1i}=0 \\ \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \displaystyle\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\,x_{mi}=0 \end{cases}$$

がなりたつとする。このとき任意の実数 $A_0, A_1, \cdots, A_m$ に対して

$$\begin{aligned}
&\sum_{i=1}^{n}\{y_i-(A_0x_{0i}+A_1x_{1i}+\cdots+A_mx_{mi})\}^2 \\
&=\sum_{i=1}^{n}\Big[\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}+\{(a_0-A_0)x_{0i}+(a_1-A_1)x_{1i}+\cdots+(a_m-A_m)x_{mi}\}\Big]^2 \\
&=\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}^2+\sum_{i=1}^{n}\{(a_0-A_0)x_{0i}+(a_1-A_1)x_{1i}+\cdots+(a_m-A_m)x_{mi}\}^2 \\
&\quad+2\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\{(a_0-A_0)x_{0i}+(a_1-A_1)x_{1i}+\cdots+(a_m-A_m)x_{mi}\}
\end{aligned}$$

となる。この最後の項は $(*)$ より

$$\begin{aligned}
&2\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\{(a_0-A_0)x_{0i}+(a_1-A_1)x_{1i}+\cdots+(a_m-A_m)x_{mi}\} \\
&=2(a_0-A_0)\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\,x_{0i} \\
&\quad+2(a_1-A_1)\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\,x_{1i} \\
&\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad+2(a_m-A_m)\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}\,x_{mi} \\
&=0
\end{aligned}$$

となるので

$$\begin{aligned}
&\sum_{i=1}^{n}\{y_i-(A_0x_{0i}+A_1x_{1i}+\cdots+A_mx_{mi})\}^2 \\
&=\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}^2+\sum_{i=1}^{n}\{(a_0-A_0)x_{0i}+(a_1-A_1)x_{1i}+\cdots+(a_m-A_m)x_{mi}\}^2 \\
&\geqq\sum_{i=1}^{n}\{y_i-(a_0x_{0i}+a_1x_{1i}+\cdots+a_mx_{mi})\}^2
\end{aligned}$$

(証明終)

＜最小2乗法＞

多項式回帰や重回帰分析で実際に用いられる最小2乗法は次の定理である。

＜定理＞

n 個の $m+1$ 変数データ

$$(x_{11}, x_{21}, \cdots, x_{m1}, y_1), \quad (x_{12}, x_{22}, \cdots, x_{m2}, y_2), \quad \cdots, \quad (x_{1n}, x_{2n}, \cdots, x_{mn}, y_n)$$

に対し

$$\sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_m x_{mi})\}^2$$

が最小になるのは $a_0, a_1, \cdots, a_m$ が次の連立方程式をみたすときである。

$$(*) \begin{cases} \displaystyle\sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_m x_{mi})\} = 0 \\ \displaystyle\sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_m x_{mi})\} x_{1i} = 0 \\ \displaystyle\sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_m x_{mi})\} x_{2i} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \displaystyle\sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_m x_{mi})\} x_{mi} = 0 \end{cases}$$

この定理は31ページの定理で $x_{0i} = 1 \ \ (i = 1, \cdots, n)$ の場合である。

正規方程式 $(*)$ を $a_0, a_1, a_2, \cdots, a_m$ の連立一次方程式の形にすると

$$(*)' \begin{cases} n a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i}\right) a_1 + \left(\displaystyle\sum_{i=1}^{n} x_{2i}\right) a_2 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi}\right) a_m = \displaystyle\sum_{i=1}^{n} y_i \\ \left(\displaystyle\sum_{i=1}^{n} x_{1i}\right) a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i} x_{1i}\right) a_1 + \left(\displaystyle\sum_{i=1}^{n} x_{2i} x_{1i}\right) a_2 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi} x_{1i}\right) a_m = \displaystyle\sum_{i=1}^{n} y_i x_{1i} \\ \left(\displaystyle\sum_{i=1}^{n} x_{2i}\right) a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i} x_{2i}\right) a_1 + \left(\displaystyle\sum_{i=1}^{n} x_{2i} x_{2i}\right) a_2 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi} x_{2i}\right) a_m = \displaystyle\sum_{i=1}^{n} y_i x_{2i} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \left(\displaystyle\sum_{i=1}^{n} x_{mi}\right) a_0 + \left(\displaystyle\sum_{i=1}^{n} x_{1i} x_{mi}\right) a_1 + \left(\displaystyle\sum_{i=1}^{n} x_{2i} x_{mi}\right) a_2 + \cdots + \left(\displaystyle\sum_{i=1}^{n} x_{mi} x_{mi}\right) a_m = \displaystyle\sum_{i=1}^{n} y_i x_{mi} \end{cases}$$

となる。

＜正規方程式＞

前ページの正規方程式 $(*)'$ を実際に計算するときはコンピューターを使う。最近は $Mathematica$ 等で行列の計算ができるようになったので、$(*)'$ を行列で表示しよう。

$$(*)' \quad \begin{pmatrix} n & \left(\sum_{i=1}^n x_{1i}\right) & \left(\sum_{i=1}^n x_{2i}\right) & \cdots & \left(\sum_{i=1}^n x_{mi}\right) \\ \left(\sum_{i=1}^n x_{1i}\right) & \left(\sum_{i=1}^n x_{1i}x_{1i}\right) & \left(\sum_{i=1}^n x_{2i}x_{1i}\right) & \cdots & \left(\sum_{i=1}^n x_{mi}x_{1i}\right) \\ \left(\sum_{i=1}^n x_{2i}\right) & \left(\sum_{i=1}^n x_{1i}x_{2i}\right) & \left(\sum_{i=1}^n x_{2i}x_{2i}\right) & \cdots & \left(\sum_{i=1}^n x_{mi}x_{2i}\right) \\ \vdots & \vdots & \vdots & \vdots & \\ \left(\sum_{i=1}^n x_{mi}\right) & \left(\sum_{i=1}^n x_{1i}x_{mi}\right) & \left(\sum_{i=1}^n x_{2i}x_{mi}\right) & \cdots & \left(\sum_{i=1}^n x_{mi}x_{mi}\right) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{1i} \\ \sum_{i=1}^n y_i x_{2i} \\ \vdots \\ \sum_{i=1}^n y_i x_{mi} \end{pmatrix}$$

ここで

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix}, \quad \mathbf{X}^t = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

とおくと

$$\mathbf{X}^t\mathbf{Y} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{1i} \\ \vdots \\ \sum_{i=1}^n y_i x_{mi} \end{pmatrix} = (*)' \text{ の右辺}$$

$$\mathbf{X}^t\mathbf{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix} = \begin{pmatrix} n & (\Sigma x_{1i}) & (\Sigma x_{2i}) & \cdots & (\Sigma x_{mi}) \\ (\Sigma x_{1i}) & (\Sigma x_{1i}x_{1i}) & (\Sigma x_{2i}x_{1i}) & \cdots & (\Sigma x_{mi}x_{1i}) \\ (\Sigma x_{2i}) & (\Sigma x_{1i}x_{2i}) & (\Sigma x_{2i}x_{2i}) & \cdots & (\Sigma x_{mi}x_{2i}) \\ \vdots & \vdots & \vdots & & \vdots \\ (\Sigma x_{mi}) & (\Sigma x_{1i}x_{mi}) & (\Sigma x_{2i}x_{mi}) & \cdots & (\Sigma x_{mi}x_{mi}) \end{pmatrix}$$

であるから $(*)'$ を行列で表すと

$$\mathbf{X}^t\mathbf{X}\mathbf{A} = \mathbf{X}^t\mathbf{Y} \qquad \text{(正規方程式)}$$

となる。そこでもし $\mathbf{X}^t\mathbf{X}$ の逆行列 $(\mathbf{X}^t\mathbf{X})^{-1}$ が存在すれば

$$\mathbf{A} = (\mathbf{X}^t\mathbf{X})^{-1}\ \mathbf{X}^t\mathbf{Y}$$

で $\mathbf{A}$ が求まる。

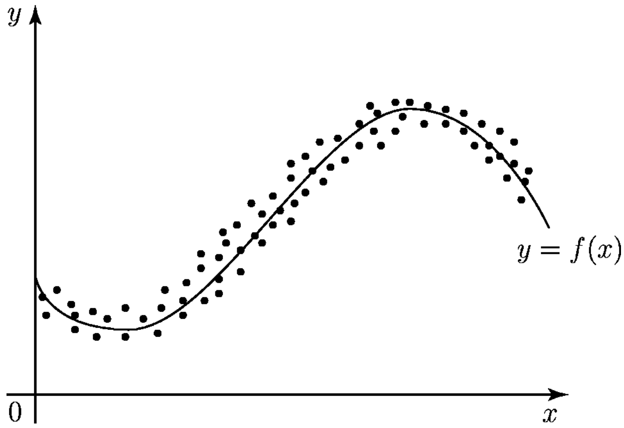
＜多項式回帰＞

n 個の 2 変数データ

$(x_1, y_1), \cdots, (x_n, y_n)$

の散布図が右図のような場合は、このデータの x 成分と y 成分の関係を直線であてはめることはできない。このような場合はある関数 $f(x)$ に対し、

$$\sum_{i=1}^{n} \{y_i - f(x_i)\}^2$$



が最小になるように関数 $f(x)$ を求めたい。しかし $f(x)$ に何らかの制限がないと $f(x)$ は求まらない。ここでは $f(x)$ を m 次の多項式

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

として、

$$\sum_{i=1}^{n} \{y_i - f(x_i)\}^2 = \sum_{i=1}^{n} \left\{y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_m x_i^m)\right\}^2$$

の値を最小にするように係数 $a_0, a_1, \cdots, a_m$ を決める。

33 ページの定理で

$$x_{1i} = x_i, \quad x_{2i} = x_i^2, \quad x_{3i} = x_i^3, \quad \cdots, \quad x_{mi} = x_i^m$$

の場合に相当する。従って 34 ページから

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

とおくと正規方程式は

$$\mathbf{X}^t \mathbf{X} \mathbf{A} = \mathbf{X}^t \mathbf{Y} \quad \text{(正規方程式)}$$

となり、この解 $\mathbf{A}$ を求めれば係数 $a_0, a_1, \cdots, a_m$ が求まる。この $a_0, a_1, \cdots, a_m$ に対する多項式 $f(x)$ を**回帰多項式**という。この回帰多項式 $f(x)$ に対し $\hat{y}_i = f(x_i)$ とおけば残差平方和は

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \Big(y_i - f(x_i)\Big)^2 \quad \text{(残差平方和)}$$

である。30 ページと同様に

$$\boxed{R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad \text{(決定係数)}$$

を決定係数といい、R^2 を回帰多項式のあてはまりの尺度とする。

< 重回帰 >

例 親の身長が子供の身長にどの程度影響を与えるかを統計的に調べたい。
父親、母親および息子の身長を n 家族分計ったデータが

$$(x_{11}, x_{21}, y_1), \quad (x_{12}, x_{22}, y_2), \quad \cdots, \quad (x_{1n}, x_{2n}, y_n)$$

とする。x_{1i} は i 番目の家族の父親の身長、x_{2i} は i 番目の家族の母親の身長、y_i は i 番目の家族の息子の身長とする。

父親の身長 x_1 と母親の身長 x_2 に対し、息子の身長 y が

$$y = a_0 + a_1 x_1 + a_2 x_2 \qquad (a_0, a_1, a_2 \text{ は定数})$$

という一次式で予測することを考える。このとき i 番目の家族の予測値は

$$\hat{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} \qquad (\text{予測値})$$

であり、予測値 $\hat{y}_i$ と実際の息子の身長 y_i との誤差は

$$y_i - \hat{y}_i = y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i}) \qquad (\text{誤差})$$

であり、誤差をできるだけ少なくするために最小 2 乗法を使う。すなわち

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i})\}^2$$

を最小にするような a_0, a_1, a_2 を求める。33, 34 ページ (で $m = 2$ の場合) より

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

に対し

$$\mathbf{X}^t \mathbf{X} \mathbf{A} = \mathbf{X}^t \mathbf{Y} \qquad (\text{正規方程式})$$

をみたす $\mathbf{A}$ を求め、それによって a_0, a_1, a_2 が求まる。この a_0, a_1, a_2 に対する予測式

$$\boxed{y = a_0 + a_1 x_1 + a_2 x_2} \qquad (\text{重回帰式})$$

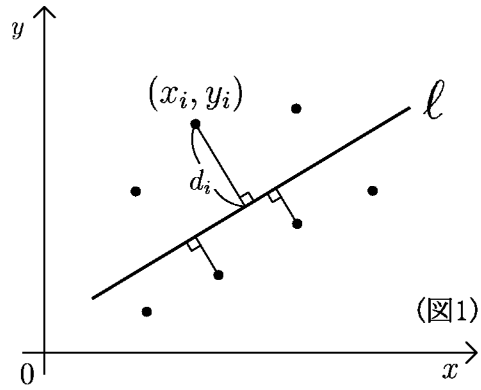
を**重回帰式**という。(x_1, x_2) を**説明変数**、y を**目的変数**という。
重回帰式の (実際のデータに対する) あてはまりの尺度は前ページと同様に

$$\text{決定係数} \ : \ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (\hat{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i})$$

で判断する。

一般に説明変数が m 個の場合は 33, 34 ページの最小 2 乗法を用いる。

＜直交回帰直線＞



n 個の 2 変数データ

$(x_1, y_1), (x_2, y_2), \cdots, (x_i, y_i), \cdots, (x_n, y_n)$

を右図のように散布図にする。このとき第 i 番目の点 (x_i, y_i) と直線 ℓ との距離を d_i とする。この距離の 2 乗の和

$$\sum_{i=1}^{n} {d_i}^2$$

を最小にするような直線 ℓ をこのデータの**直交回帰直線**という。

このデータの統計量

$$S_{xx} = \frac{1}{n}\sum_{i=i}^{n}(x_i - \overline{x})^2, \quad S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2, \quad S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

に対し、直交回帰直線 ℓ の方程式は以下のように定まる。

〔I〕＜ $S_{xy} \neq 0$ のとき＞

$$\ell : \boxed{y = \frac{S_{yy} - S_{xx} + \sqrt{D}}{2S_{xy}}(x - \overline{x}) + \overline{y}} \text{ (直交回帰直線)}$$

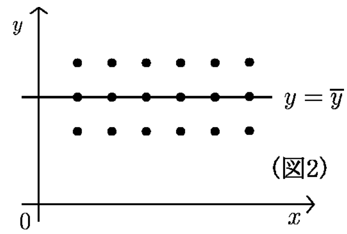
ただし $D = (S_{xx} - S_{yy})^2 + 4(S_{xy})^2$ である。直交回帰直線は $(\overline{x}, \overline{y})$ を必ず通る。

〔II〕＜ $S_{xy} = 0$ のとき＞

(1) $S_{xx} > S_{yy}$ のとき直交回帰直線は

$$\ell : \boxed{y = \overline{y}}$$

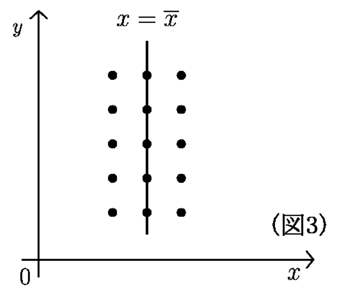
になる (図 2)。



(2) $S_{xx} < S_{yy}$ のとき直交回帰直線は

$$\ell : \boxed{x = \overline{x}}$$

になる (図 3)。

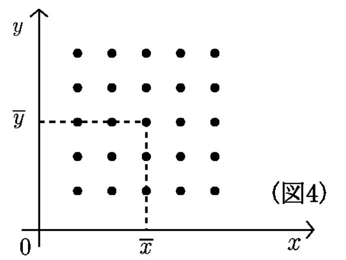


(3) $S_{xx} = S_{yy}$ のとき直交回帰直線は決まらない (図 4)。

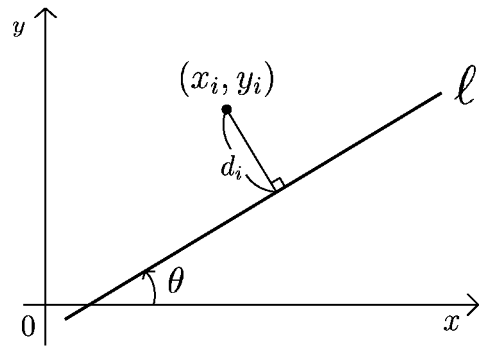
実は散布図の中心 $(\overline{x}, \overline{y})$ を通る直線に対し、常に

$$\sum_{i=1}^{n} {d_i}^2 = \frac{n}{2}\big(S_{xx} + S_{yy}\big) = nS_{xx}$$

が成り立つ。



＜直交回帰直線の導出１＞



前ページの直交回帰直線 ℓ の式を導く。
右図のように x 軸からの角度を θ とすると、
傾きは $\tan\theta$ となる。今 ℓ の式を

$$\ell : y = \tan\theta(x - \overline{x}) + \overline{y} + k$$

とおくと ℓ は

$$\ell : \sin\theta(x - \overline{x}) - \cos\theta(y - \overline{y}) + k\cos\theta = 0$$

となる。点 (x_i, y_i) と直線 ℓ との距離 d_i は

$$d_i = \frac{|\sin\theta(x_i - \overline{x}) - \cos\theta(y_i - \overline{y}) + k\cos\theta|}{\sqrt{\sin^2\theta + \cos^2\theta}} = |\sin\theta(x_i - \overline{x}) - \cos\theta(y_i - \overline{y}) + k\cos\theta|$$

より

$$\sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} \Big\{\sin\theta(x_i - \overline{x}) - \cos\theta(y_i - \overline{y}) + k\cos\theta\Big\}^2$$

$$= \sin^2\theta \sum_{i=1}^{n}(x_i - \overline{x})^2 + \cos^2\theta \sum_{i=1}^{n}(y_i - \overline{y})^2 - 2\sin\theta\cos\theta \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) + nk^2\cos^2\theta$$

となる。ここで $\displaystyle\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, $\displaystyle\sum_{i=1}^{n}(y_i - \overline{y}) = 0$ を使った。S_{xx}, S_{yy}, S_{xy} の定義より

$$\sum_{i=1}^{n} {d_i}^2 = n\Big\{\sin^2\theta S_{xx} + \cos^2\theta S_{yy} - 2\sin\theta\cos\theta S_{xy}\Big\} + nk^2\cos^2\theta$$

となる。従って k に関して $\displaystyle\sum_{i=1}^{n} {d_i}^2$ は $k = 0$ のときが最小となるので、今後 $k = 0$ とする。

半角の公式より

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} {d_i}^2 &= \frac{1 - \cos(2\theta)}{2}S_{xx} + \frac{1 + \cos(2\theta)}{2}S_{yy} - \sin(2\theta)S_{xy} \\ &= \frac{S_{xx} + S_{yy}}{2} - \left\{\frac{S_{xx} - S_{yy}}{2}\cos(2\theta) + S_{xy}\sin(2\theta)\right\} \\ &= \frac{S_{xx} + S_{yy}}{2} - r\sin(2\theta + \alpha)\end{aligned}$$

となる。ただし

$$\sin\alpha = \frac{S_{xx} - S_{yy}}{2r}, \quad \cos\alpha = \frac{S_{xy}}{r}, \quad r = \sqrt{\left(\frac{S_{xx} - S_{yy}}{2}\right)^2 + (S_{xy})^2}$$

である。従って $\displaystyle\sum_{i=1}^{n} {d_i}^2$ は $\sin(2\theta + \alpha) = 1$ のとき $\left(2\theta + \alpha = \dfrac{\pi}{2}\text{のとき}\right)$ 最小になる。

そのとき $(2\theta = \dfrac{\pi}{2} - \alpha\text{のとき})$ 直交回帰直線 ℓ の傾き ($\tan\theta$の値) を次ページで求める。

< 直交回帰直線の導出 2 >

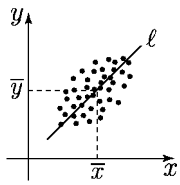
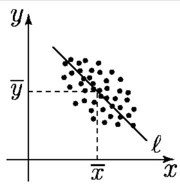
前ページより $\sin\alpha = \dfrac{S_{xx}-S_{yy}}{2r}$, $\cos\alpha = \dfrac{S_{xy}}{r}$, $r = \sqrt{\left(\dfrac{S_{xx}-S_{yy}}{2}\right)^2 + (S_{xy})^2}$, $2\theta = \dfrac{\pi}{2} - \alpha$ から

$$\cos(2\theta) = \cos\left(\frac{\pi}{2}-\alpha\right) = \sin\alpha = \frac{S_{xx}-S_{yy}}{2r}\ ,\quad \sin(2\theta) = \sin\left(\frac{\pi}{2}-\alpha\right) = \cos\alpha = \frac{S_{xy}}{r}\ ,$$

$\tan(2\theta) = \dfrac{\sin(2\theta)}{\cos(2\theta)} = \dfrac{2\tan\theta}{1-\tan^2\theta}$ より $(\tan\theta)^2 + 2\left(\dfrac{\cos(2\theta)}{\sin(2\theta)}\right)\tan\theta - 1 = 0$ から

$$\tan\theta = -\frac{\cos(2\theta)}{\sin(2\theta)} \pm \sqrt{\left(\frac{\cos(2\theta)}{\sin(2\theta)}\right)^2 + 1} = \frac{-\cos(2\theta)\pm 1}{\sin(2\theta)} = \frac{S_{yy}-S_{xx}\pm 2r}{2S_{xy}}$$

となる。ここで $-\dfrac{\pi}{2} < \theta \leqq \dfrac{\pi}{2}$, $-\dfrac{\pi}{2} \leqq \alpha < \dfrac{3\pi}{2}$ という範囲にとると、次の表のようになる。

S_{xy}	α	$\theta = \frac{\pi}{4} - \frac{\alpha}{2}$	$2S_{xy}\tan\theta = S_{yy} - S_{xx} \pm 2r$	$\tan\theta$	散布図
$S_{xy} > 0$ $(\cos\alpha > 0)$	$-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$	$0 < \theta < \frac{\pi}{2}$ $(\tan\theta > 0)$	$2S_{xy}\tan\theta = S_{yy} - S_{xx} \pm 2r > 0$	$\tan\theta = \dfrac{S_{yy} - S_{xx} + 2r}{2S_{xy}}$	
$S_{xy} < 0$ $(\cos\alpha < 0)$	$\frac{\pi}{2} < \alpha < \frac{3\pi}{2}$	$-\frac{\pi}{2} < \theta < 0$ $(\tan\theta < 0)$	$2S_{xy}\tan\theta = S_{yy} - S_{xx} \pm 2r > 0$	$\tan\theta = \dfrac{S_{yy} - S_{xx} + 2r}{2S_{xy}}$	

ここで $2r = \sqrt{(S_{xx}-S_{yy})^2 + 4(S_{xy})^2} \geqq |S_{xx} - S_{yy}|$ より

$$S_{yy} - S_{xx} + 2r \geqq 0 \qquad , \qquad S_{yy} - S_{xx} - 2r \leqq 0$$

が常になりたつので上の表のように $\tan\theta$ が決まる。

次に $S_{xy} = 0$ のときを考える。前ページより $k = 0$ だから直交回帰直線は常にデータの中心 ($\bar{x}$, $\bar{y}$) を通るので次の表のようになる。

S_{xy}	$S_{xx} - S_{yy}$	α	$\theta = \frac{\pi}{4} - \frac{\alpha}{2}$	直交回帰直線	散布図
$S_{xy} = 0$ $(\cos\alpha = 0)$	$S_{xx} - S_{yy} > 0$ $(\sin\alpha > 0)$	$\alpha = \frac{\pi}{2}$	$\theta = 0$	$y = \bar{y}$ (x 軸に平行)	37 ページ図 2
	$S_{xx} - S_{yy} < 0$ $(\sin\alpha < 0)$	$\alpha = -\frac{\pi}{2}$	$\theta = \frac{\pi}{2}$	$x = \bar{x}$ (y 軸に平行)	37 ページ図 3

最後に $S_{xy} = 0$, $S_{xx} - S_{yy} = 0$ のときは前ページの式より

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \frac{S_{xx}+S_{yy}}{2}$$

となり、θ に無関係になるので、θ は定まらない。以上をまとめると 37 ページの結果になる。

< 直交回帰直線の応用 >

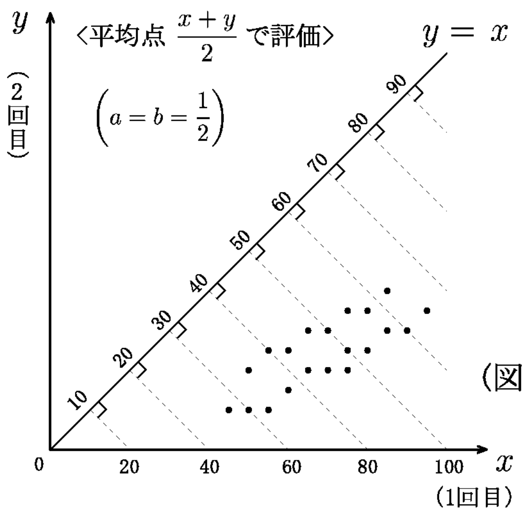
例 20 人のクラスで数学のテストを 2 回やった結果が次の表である。

1 回目の点 x	45	50	50	55	55	60	60	65	65	70	70	75	75	75	80	80	85	85	90	95
2 回目の点 y	10	10	20	10	25	15	25	20	30	20	30	20	25	35	25	35	30	40	30	35

両方とも 100 点満点であるが、2 回目のテストは難しかったのであまり点差が開かない。通常は両方の点の平均 $\dfrac{x+y}{2}$ で数学の学力を評価するが、この場合同じ比率では学力を正確に評価できない。1 回目の方が学力差が大きいので、1 回目の方をより高く評価すべきである。そこで $0<a<1,\ 0<b<1\ (a+b=1)$ に対して評価点を

$$z = ax + by \quad (\text{評価点})$$

として、a,b を以下のように決定する。



(図1)

[平均で評価する場合は散布図上に直線 $y=x$ を描き、この直線上に図 1 のような目もりを入れ、散布図上の各点から $y=x$ におろした垂線の目もりを読むと平均点が分かる。]

この [ ] の方法を直交回帰直線に適用する。まず図 2 のように直交回帰直線をひき、同じ傾きで原点を通る直線 $y=(\tan\theta)x$ 上に目もりを入れる。この目もりの決め方を次のように定める。図 3 のように各点 (x_i,y_i) から直線 $y=(\tan\theta)x$ におろした垂線の足を点 P とする。点 P の目もりを z_i とすると、z_i は (x_i,y_i) の評価点

$$z_i = ax_i + by_i \quad (\text{点 P の目もり})$$

である。一方 z_i は OP の長さに比例する。図 3 より

$$\mathrm{OP} = k\cos\theta = (x_i + y_i\tan\theta)\cos\theta = x_i\cos\theta + y_i\sin\theta$$

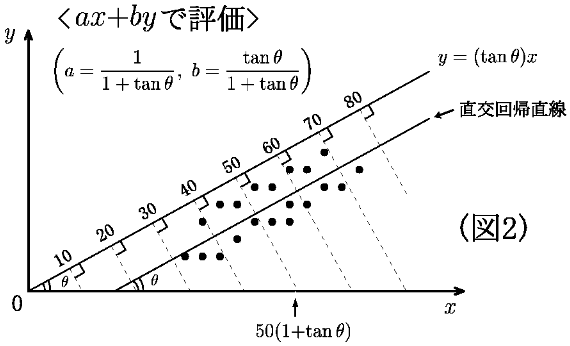
であるから

$$a : b = \cos\theta : \sin\theta$$

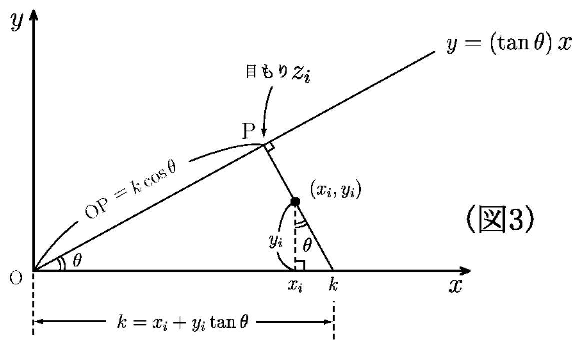
となる。$a+b=1$ から

$$a = \frac{\cos\theta}{\cos\theta+\sin\theta} = \frac{1}{1+\tan\theta}$$

$$b = \frac{\sin\theta}{\cos\theta+\sin\theta} = \frac{\tan\theta}{1+\tan\theta}$$



(図2)



(図3)

が求まる。目もり z_i と x 軸上の点 $(k,0)$ との関係は $z_i = ak$ より $k = z_i\times(1+\tan\theta)$ である。

(注 1) より正確に評価するためには 1 回目の偏差値 $\left(\frac{x_i-\overline{x}}{\sqrt{S_{xx}}}\times 10+50\right)$ と 2 回目の偏差値 $\left(\frac{y_i-\overline{y}}{\sqrt{S_{yy}}}\times 10+50\right)$ の散布図の直交回帰直線を求め、例のように a,b を求めて評価する。

(注 2) このように 2 つの変量を 1 つにまとめる方法を一般化したものを「**主成分分析**」という。回帰分析や主成分分析については「多変量解析」の本に詳しく説明してある。