相関と最小２乗法

静岡大学工学部機械工学科
岡本正芳

相関

標本は2対のペアー(x_i, y_i)で与えられる場合も頻繁にある。

例えば，人における身長と体重，時間と降水量，温度と金属の膨張率，位置と速度など様々なものが存在する。

x_i と y_i の間に関連性があるかどうかを議論する。

相関評価の手順

I. データーの収集
II. データーを散布図によりグラフ化
III. 相関係数の算出
IV. 最適関数の同定

評価データー

表 6.1 解析例データー

x_i	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(3)}$	$y_i^{(4)}$	$y_i^{(5)}$
0.2	0.640	0.501	2.624	2.875	2.715
0.4	0.732	0.798	0.678	2.781	1.963
0.6	0.900	0.572	0.533	2.112	1.480
0.8	1.037	0.657	0.678	2.031	1.152
1.0	1.126	1.041	1.349	1.461	0.532
1.2	1.300	1.664	1.502	1.659	0.342
1.4	1.467	1.518	0.312	1.831	0.417
1.6	1.546	1.625	3.087	1.443	0.169
1.8	1.699	1.646	0.193	0.951	0.137
2.0	1.802	1.949	2.114	1.243	0.087
2.2	2.015	1.884	2.184	0.765	0.077
2.4	2.208	1.702	2.946	0.812	0.031
2.6	2.233	1.885	4.060	1.136	0.025
2.8	2.352	2.153	2.079	0.100	0.021

グラフによる可視化

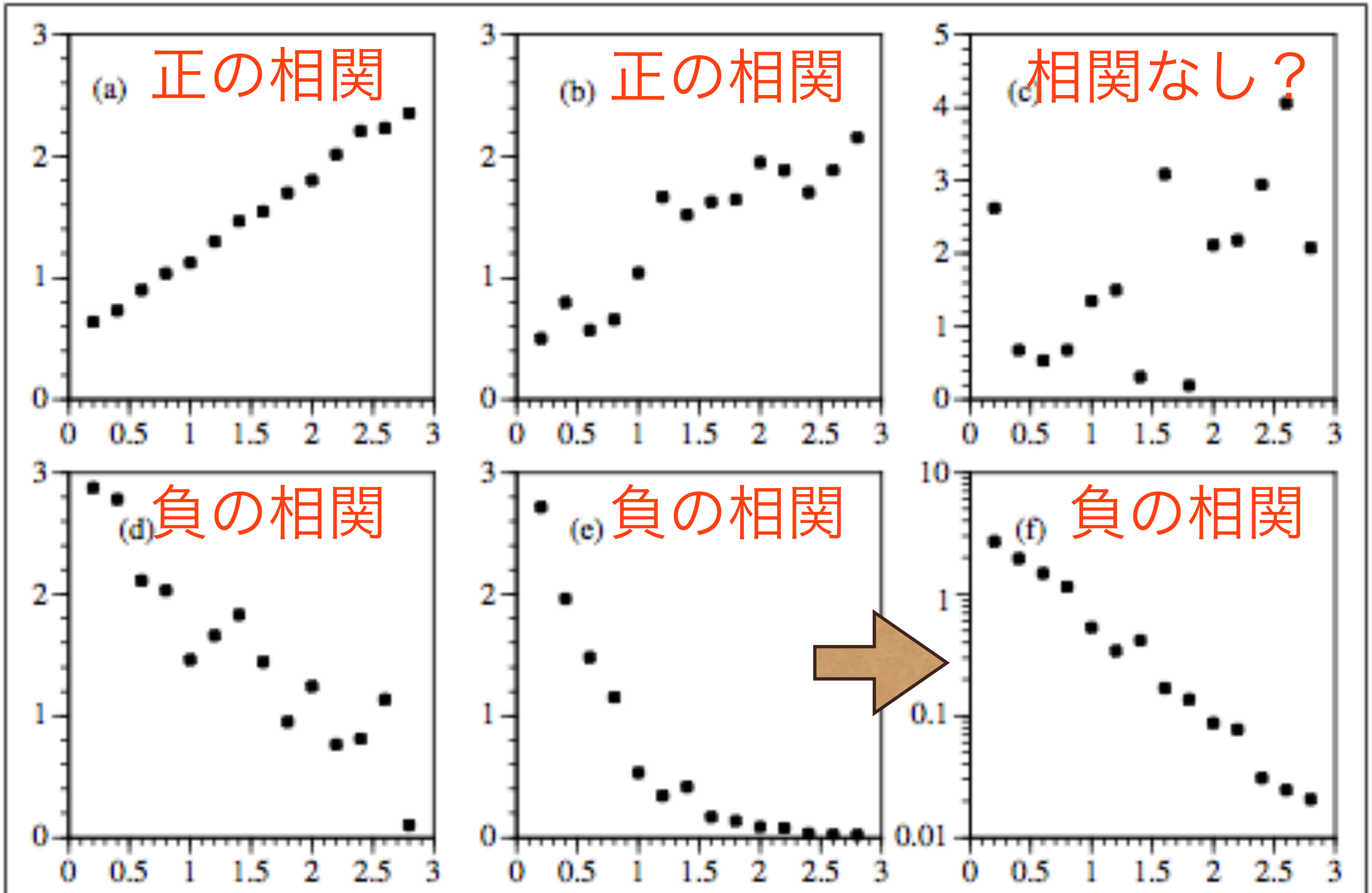

図 6.1　解析例データーの散布図

相関係数

- 相関係数の定義式

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

- rが+1に近い：正の相関
- rが0付近　　：相関なし
- rが−1に近い：負の相関

xとyの関係性を調べる

最小２乗法（一般形）

n個の標本セット(x_i, y_i) があるとき、その関係性がk+1個のパラメーターセット$(a_0, a_1, ..., a_k)$を用いた関数

$$y = G\left(x; a_0, a_1, \cdots, a_k\right)$$

で近似できるとする。真値y_iと推定値$G(x_i)$の差の２乗和Vを以下で与える。

$$V\left(a_0, a_1, \cdots, a_k\right) = \sum_{i=1}^{n}\left(y_i - G\left(x_i; a_0, a_1, \cdots, a_k\right)\right)^2$$

最小２乗法（一般形）

Vが最小となるようにパラメーターセット$(a_0, a_1, ..., a_k)$を決定することで近似関数を決定する。

決定式

$$\frac{\partial V}{\partial a_j} = -2\sum_{i=1}^{n}\left(y_i - G\left(x_i; a_0, a_1, \cdots, a_k\right)\right)\frac{\partial G\left(x_i; a_0, a_1, \cdots, a_k\right)}{\partial a_j} = 0$$

$$\frac{\partial^2 V}{\partial a_j^2} = 2\sum_{i=1}^{n}\frac{\partial G\left(x_i; a_0, a_1, \cdots, a_k\right)}{\partial a_j}\frac{\partial G\left(x_i; a_0, a_1, \cdots, a_k\right)}{\partial a_j}$$

$$-2\sum_{i=1}^{n}\left(y_i - G\left(x_i; a_0, a_1, \cdots, a_k\right)\right)\frac{\partial^2 G\left(x_i; a_0, a_1, \cdots, a_k\right)}{\partial a_j^2} > 0$$

最小２乗法（多項式）

n個の標本セット(x_i, y_i)があるとき、その関係性がk+1個のパラメーターセット$(a_0, a_1, ..., a_k)$を用いた関数

$$y = \sum_{j=0}^{k} a_j x^j$$

で近似できるとする。真値y_iと推定値$\sum_{j=0}^{k} a_j x_i^j$の差の２乗和Vを以下で与える。

$$V(a_0, a_1, \cdots, a_k) = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} a_j x_i^j \right)^2$$

最小２乗法（多項式）

Vが最小となるようにパラメーターセット$(a_0, a_1, ..., a_k)$を決定することで近似関数を決定する。

決定式 $$\frac{\partial V}{\partial a_j} = -2\sum_{i=1}^{n} x_i^j \left(y_i - \sum_{j=0}^{k} a_j x_i^j \right) = 0$$

k+1本の連立１次方程式

$$\frac{\partial^2 V}{\partial a_j^2} = 2\sum_{i=1}^{n} x_i^{2j} > 0$$

最小値条件を満足

最小２乗法（線形近似）

n個の標本セット(x_i, y_i) があるとき、その関係性が2個のパラメーターセット(a_0, a_1)を用いた線形関数

$$y = a_0 + a_1 x$$

で近似できるとする。真値y_iと推定値$a_0+a_1x_i$の差の２乗和Vを以下で与える。

$$V(a_0, a_1) = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

よく利用するので覚えておくように！！

最小２乗法（線形関数）

Vが最小となるようにパラメーターセット(a_0, a_1)を決定することで近似関数を決定する。

$$\frac{\partial V(a_0, a_1)}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) = 0$$

$$\frac{\partial V(a_0, a_1)}{\partial a_1} = -2\sum_{i=1}^{n}x_i(y_i - a_0 - a_1 x_i) = 0$$

最小２乗法（線形関数）

第１式 $\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) = \sum_{i=1}^{n} y_i - a_0 \sum_{i=1}^{n} 1 - a_1 \sum_{i=1}^{n} x_i$

$$= n\bar{y} - na_0 - na_1\bar{x} = 0$$

$$\Rightarrow \quad a_0 = \bar{y} - a_1\bar{x}$$

第２式 $\sum_{i=1}^{n} x_i(y_i - a_0 - a_1 x_i) = \sum_{i=1}^{n} x_i y_i - a_0 \sum_{i=1}^{n} x_i - a_1 \sum_{i=1}^{n} x_i^2$

$$= \sum_{i=1}^{n} x_i y_i - n\bar{x} \cdot \bar{y} + na_1\bar{x}^2 - a_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \quad a_1 = \frac{S_{xy}}{S_{xx}}$$

$$\Rightarrow \quad a_0 = \bar{y} - \bar{x}\frac{S_{xy}}{S_{xx}}$$

最小点の確認

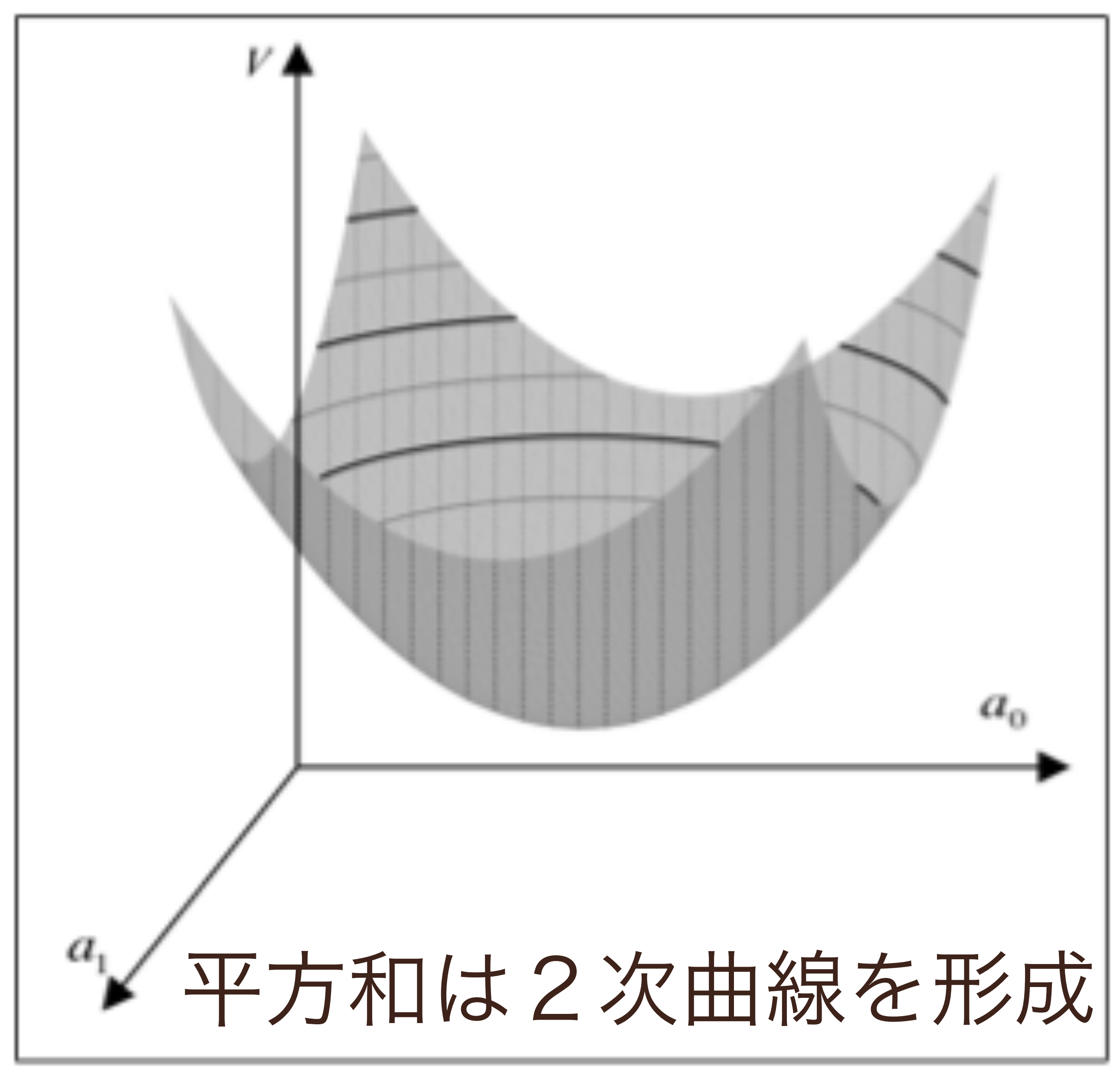

図 6.2　誤差の 2 乗和 V に関する概略図