実験データー解析概論

— 統計学に基づく「よりよい推論」のために —

三中 信宏

MINAKA Nobuhiro

独立行政法人 農業環境技術研究所 生態系計測研究領域 上席研究員［生物統計学］

東京大学大学院 農学生命科学研究科 生物・環境工学専攻 教授［生態系計測学］

東京農業大学大学院 農学研究科 客員教授［応用昆虫学］

mailto:minaka@affrc.go.jp（メール）

http://twitter.com/leeswijzer/（ツイッター）

http://cse.niaes.affrc.go.jp/minaka/（ウェブサイト）

http://d.hatena.ne.jp/leeswijzer/（ブログ）

2014 年度・東京農業大学生物応用化学科（毎週金曜第 2 限）

三中信宏〈実験データー解析概論〉：2014 年 4 月〜 2014 年 9 月

東京農業大学世田谷キャンパス 431 番教室

多変量同時確率分布

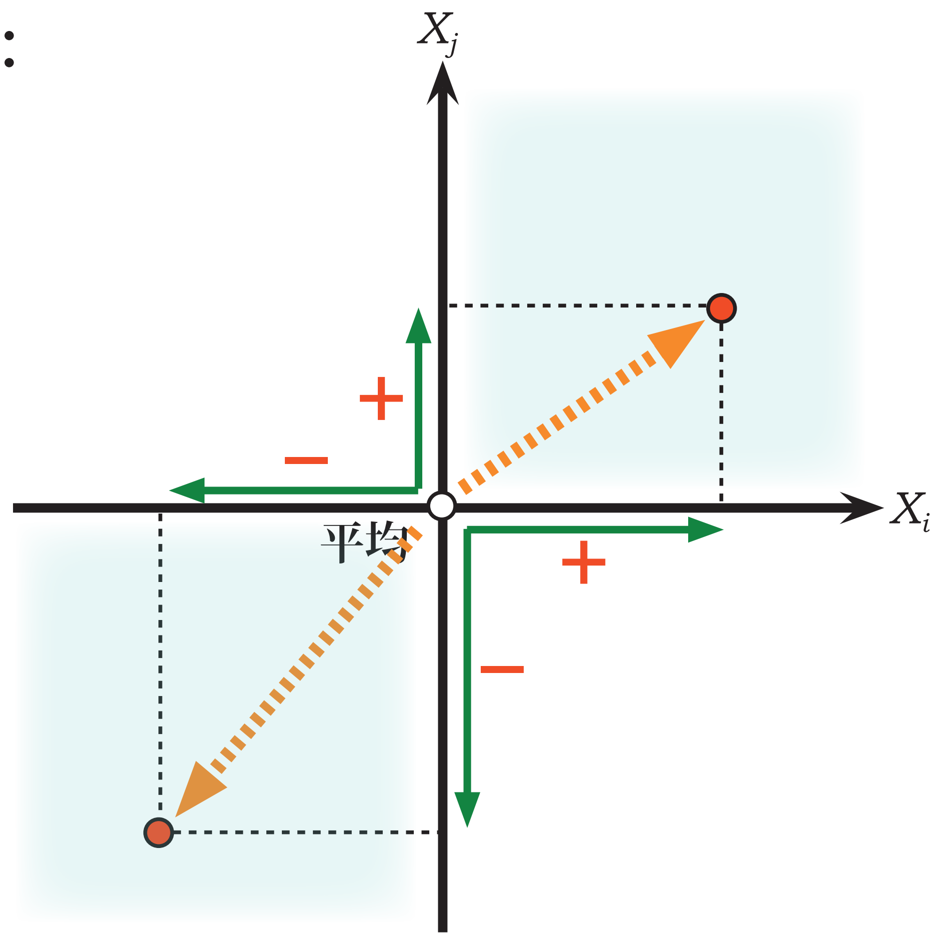
共分散を直感的に理解する

偏差が「正の共変動」をするとき：

	偏差の符号	
$X_i - E[X_i]$	$+$	$-$
$X_j - E[X_j]$	$+$	$-$
	↓	↓

$$(X_i - E[X_i])(X_j - E[X_j])$$

は「正」の値になる.



多変量同時確率分布

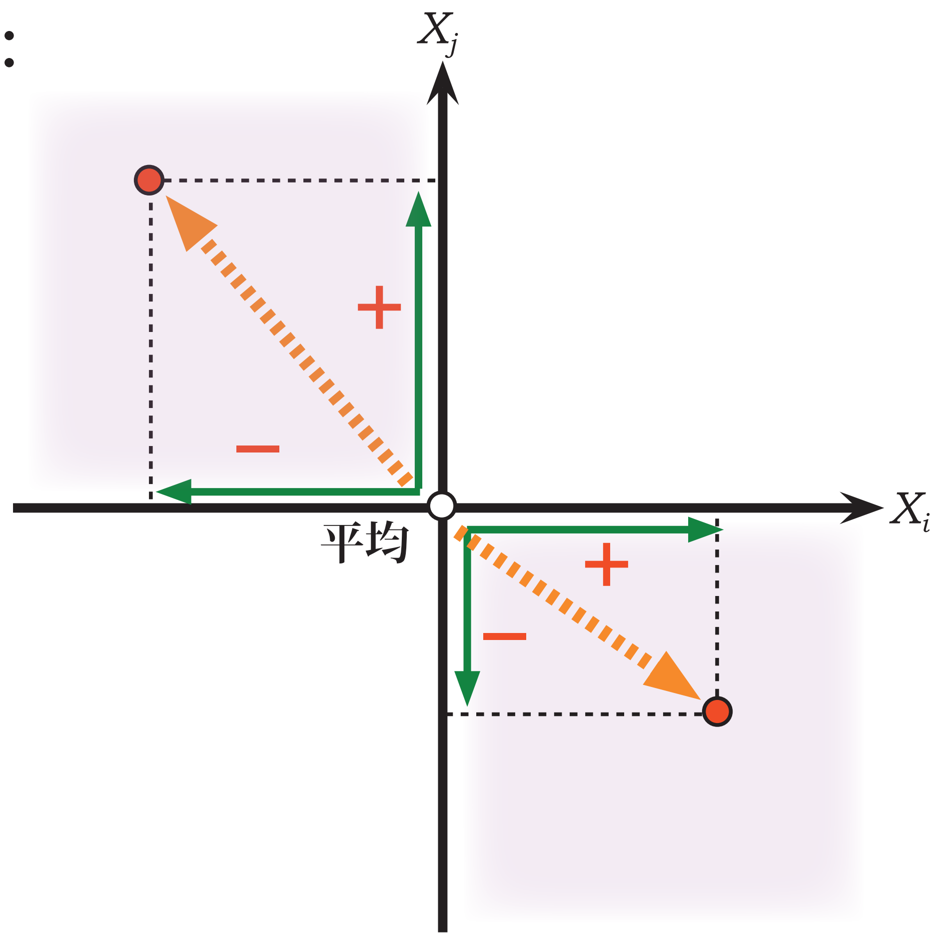
共分散を直感的に理解する

偏差が「負の共変動」をするとき：

	偏差の符号	
$X_i - E[X_i]$	+	−
$X_j - E[X_j]$	−	+
	↓	↓

$$(X_i - E[X_i])(X_j - E[X_j])$$

は「負」の値になる.



多変量同時確率分布

共分散を直感的に理解する

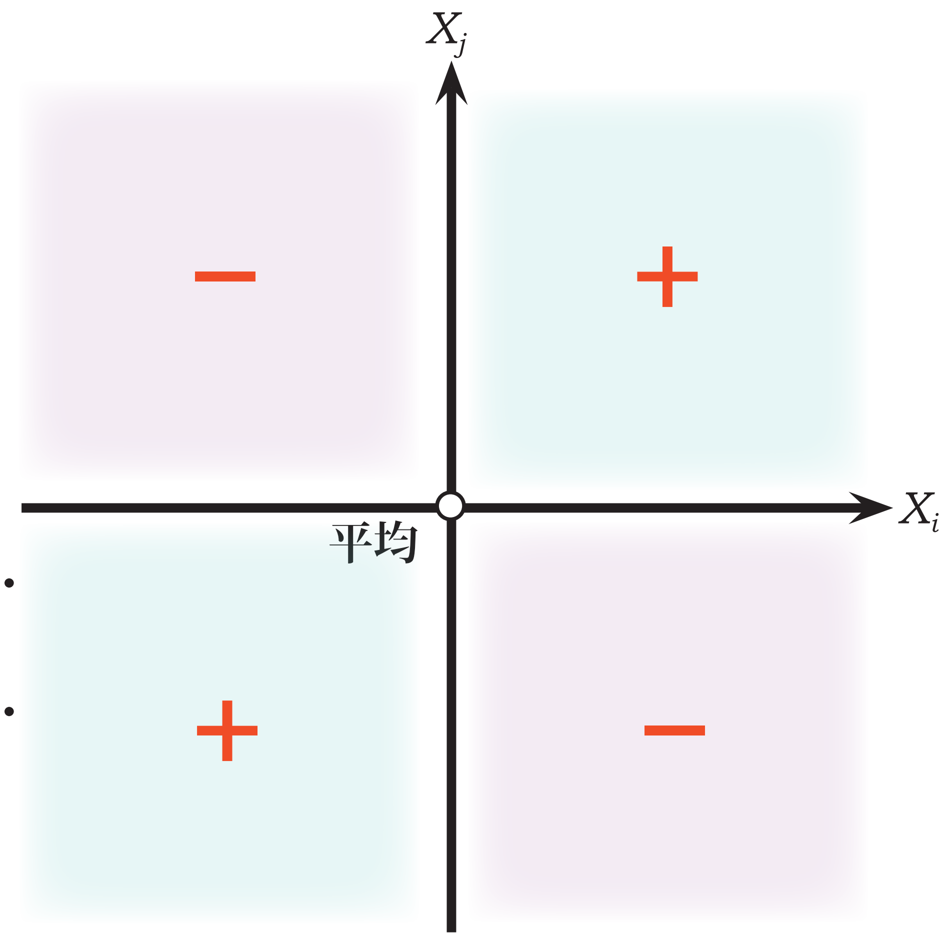
変量のペアの偏差積：

$$(X_i - E[X_i])(X_j - E[X_j])$$

の符号は,

1) 第一，第三象限で「正」になる.

2) 第二，第四象限で「負」になる.



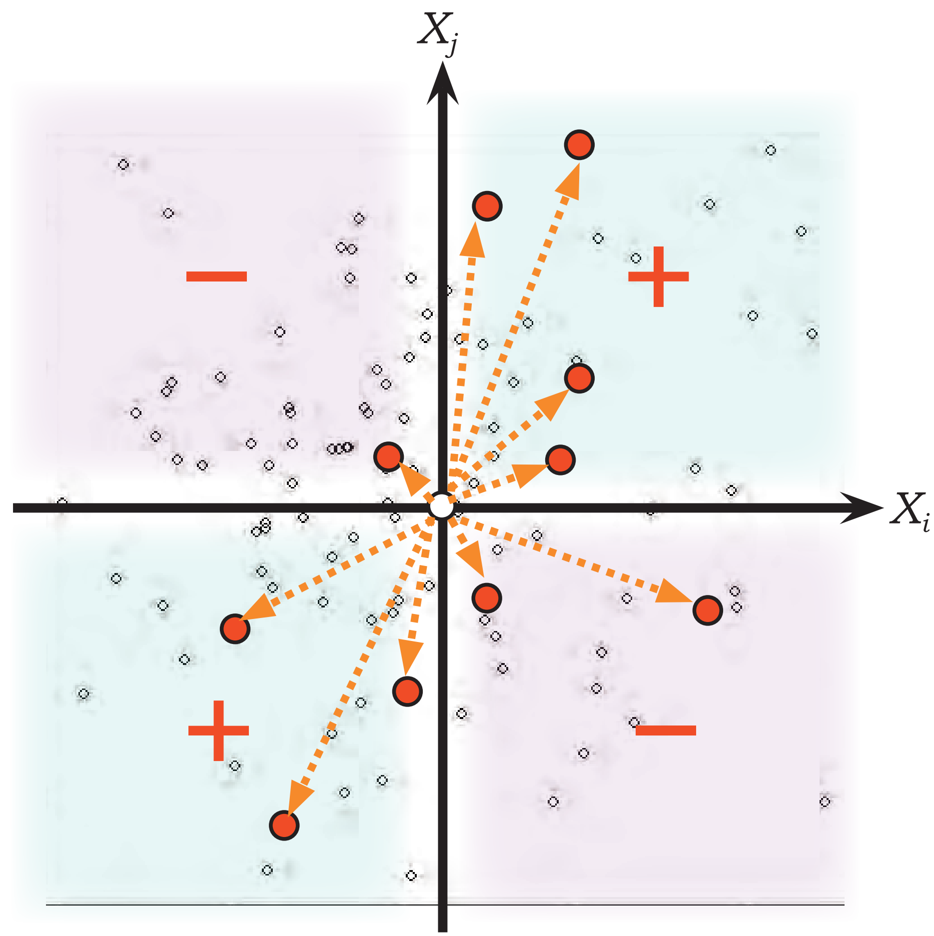
多変量同時確率分布

共分散を直感的に理解する

偏差積の期待値である共分散：

$$E[(X_i - E[X_i])(X_j - E[X_j])]$$

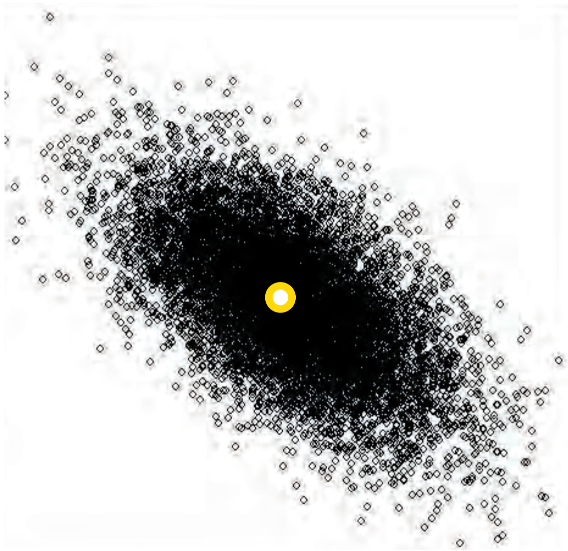
の値によって，二変量の間の共変動のパターンが明らかになる.



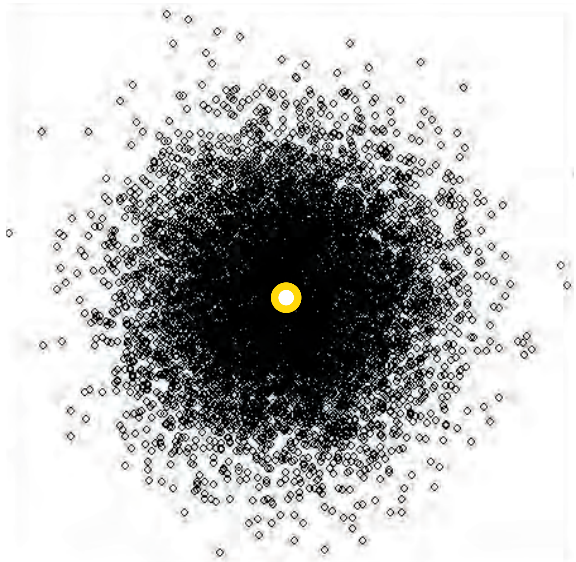
多変量同時確率分布

共分散を直感的に理解する

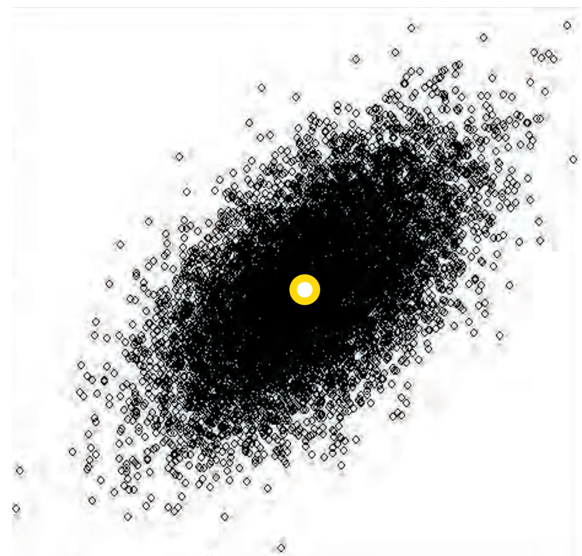
共分散が「負」



共分散が「0」



共分散が「正」



多変量同時確率分布

変量間の共分散とは何か

変量の分散（variance）と共分散（covariance）は，いずれも平均からの偏差（deviation）に関する期待値を計算している．

1)「分散」はある変量の「偏差平方」の期待値であり，平均値からの偏差の大きさすなわち変動のパラメータである．

2)「共分散」は変量間の「偏差積」の期待値であり，変量間での偏差の連動の程度すなわち共変動のパラメータである．

多変量同時確率分布

変量間の共分散とは何か

分散共分散行列の第 (i, j) 要素

$$E[(X_i - E[X_i])(X_j - E[X_j])] = \begin{cases} var[X_i] & (i = j) \\ cov[X_i, X_j] & (i \neq j) \end{cases}$$

分散

対角要素

共分散

非対角要素

偏差

各変量の値がその期待値（平均）からどれくらい離れているか

多変量同時確率分布

共分散から相関係数へ

共分散の値は，変量それ自身がもっているばらつきの程度（分散）の影響を受ける．異なる変量ペア間での共変動を相互比較するために，共分散を分散によって補正し，相関係数（correlation coefficient）を次式のように定義する．

相関係数 $$\rho = \frac{cov[X_i, X_j]}{\sqrt{var[X_i]}\sqrt{var[X_j]}}$$

相関係数は－１≦ρ≦＋１の範囲の値を取る．

多変量同時確率分布

二変量正規分布の確率密度関数

変量ベクトル $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

平均ベクトル $\boldsymbol{\mu} = \begin{pmatrix} E[X_1] \\ E[X_2] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

分散共分散行列

$$\boldsymbol{\Sigma} = \begin{pmatrix} var[X_1] & cov[X_1, X_2] \\ cov[X_2, X_1] & var[X_2] \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

多変量同時確率分布

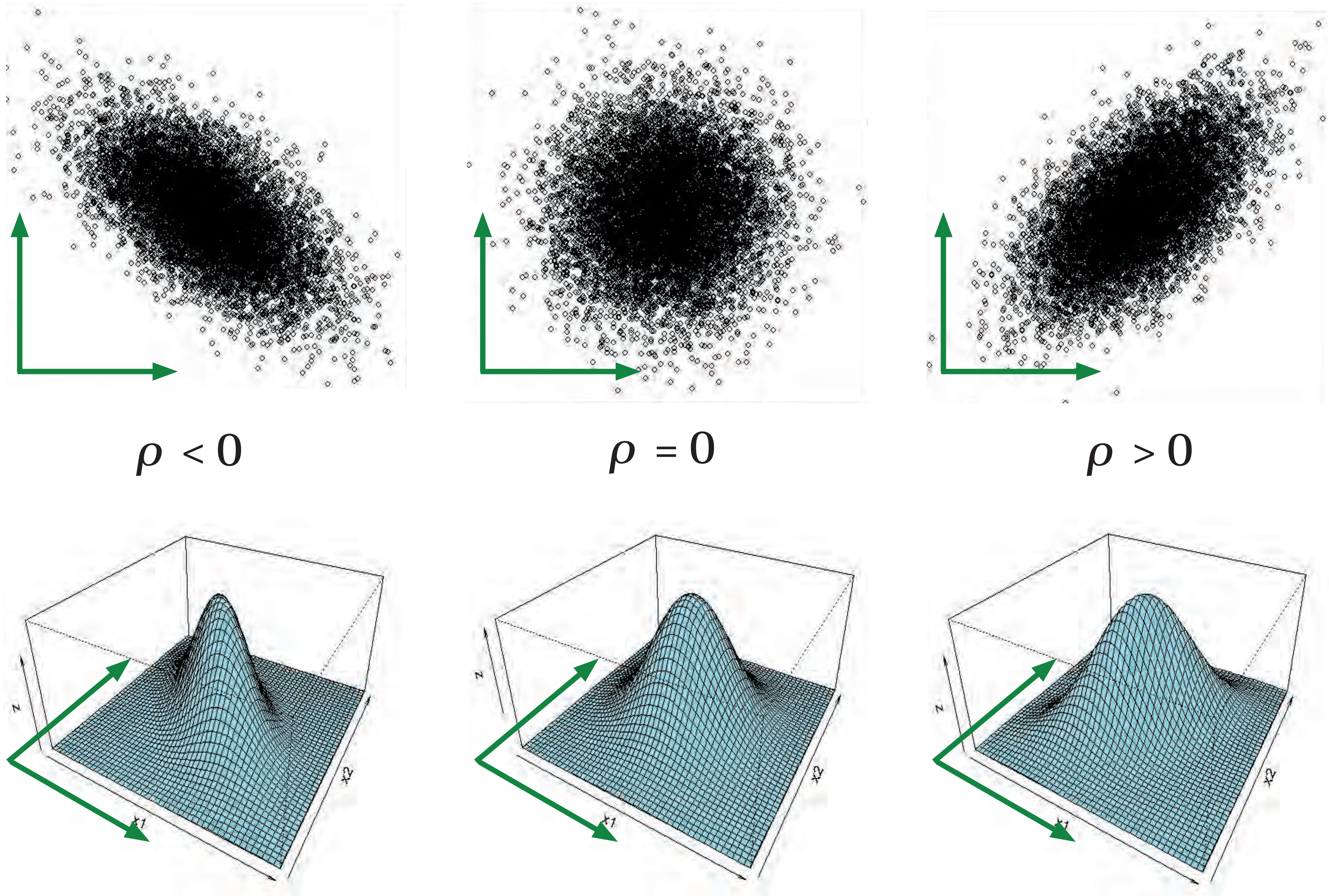
二変量正規分布の確率密度関数

確率密度関数

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right.\right.$$

$$\left.\left.-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]$$

相関係数 ρ の正負によって，変量間の共変動パターンが変化する．



ρ < 0
ρ = 0
ρ > 0
z
x1
x2
z
x1
x2
z
x1
x2

多変量同時確率分布

一変量から多変量へ

1 変量正規分布

分散（標準偏差）　　平均

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

変量

p 変量正規分布

分散共分散行列　　平均ベクトル

$$f(\boldsymbol{X}) = \frac{1}{\sqrt{2\pi}^p\sqrt{|\boldsymbol{\Sigma}|}}\exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right]$$

変量ベクトル

多変量同時確率分布

多変量正規分布をよく見ると

p 変量正規分布

$$f(\boldsymbol{X}) = \frac{1}{\sqrt{2\pi}^{p}\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right]$$

変量ベクトル：p 次元の列ベクトル

平均ベクトル：p 次元の列ベクトル

分散共分散行列：$p \times p$ 型の対称行列

多変量を「ベクトル・行列」で記述する

変量ベクトル

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

分布位置を示すパラメータ

平均ベクトル

$$\boldsymbol{\mu} = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_p] \end{pmatrix}$$

ばらつきを示すパラメータ

分散共分散行列

$$\boldsymbol{\Sigma} = \begin{pmatrix} var[X_1] & cov[X_1, X_2] & \ldots & cov[X_1, X_p] \\ cov[X_2, X_1] & var[X_2] & \ldots & cov[X_2, X_p] \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ cov[X_p, X_1] & cov[X_p, X_2] & \ldots & var[X_p] \end{pmatrix}$$

多変量同時確率分布

変量間の共分散とは何か

分散共分散行列

分散

$$\boldsymbol{\Sigma} = \begin{pmatrix} var[X_1] & cov[X_1, X_2] & \dots & cov[X_1, X_p] \\ cov[X_2, X_1] & var[X_2] & \dots & cov[X_2, X_p] \\ \dots & \dots & \dots & \dots \\ cov[X_p, X_1] & cov[X_p, X_2] & \dots & var[X_p] \end{pmatrix}$$

共分散