

確率と統計(0)

「多次元の確率分布(第7章)」

- 担当教員: 杉山 将 (計算工学専攻)
- 居室: W8E-505
- 電子メール: sugi@cs.titech.ac.jp
- 授業のウェブサイト:
<http://sugiyama-www.cs.titech.ac.jp/~sugi/>

多次元の確率分布

179

- 複数の確率変数が与えられる場合, 確率変数間の関係を調べることは重要である
- 簡単のため, 2つの確率変数 X と Y の性質を考える

- 確率変数 X と Y の同時確率分布(joint probability distribution):

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$$

$f(x, y)$: 確率変数 X と Y の同時確率密度関数
(joint probability density function):

$$f(x, y) \geq 0, \quad \int \int f(x, y) dx dy = 1$$

- 周辺確率分布(marginal probability distribution):
確率変数 X および Y 単独の確率分布 . 同時確率
密度関数から , 次式で求められる .

$$P(a \leq X \leq b) = \int_a^b \int f(x, y) dy dx = \int_a^b g(x) dx$$

$$P(c \leq Y \leq d) = \int_c^d \int f(x, y) dx dy = \int_c^d h(y) dy$$

$g(x), h(y)$: 周辺確率密度関数(marginal probability density function)

$$g(x) = \int f(x, y) dy$$

$$h(y) = \int f(x, y) dx$$

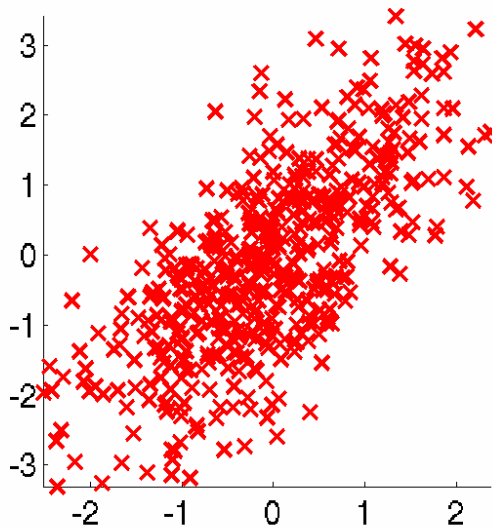
共分散

182

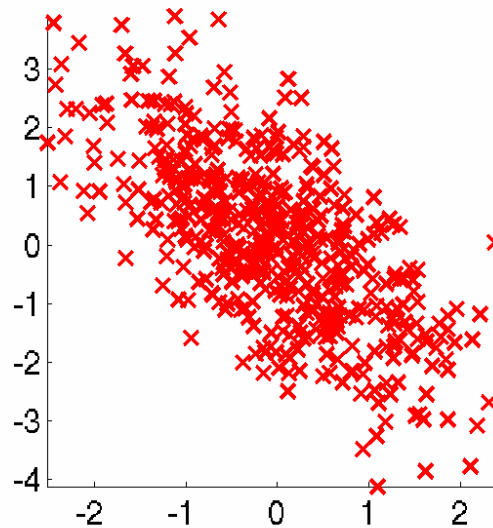
■ X と Y の共分散(covariance):

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

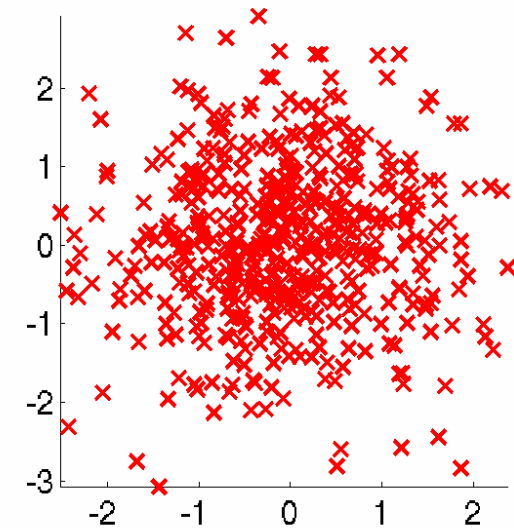
- $Cov(X, Y) > 0$ のとき, X と Y の増減は同傾向
- $Cov(X, Y) < 0$ のとき, X と Y の増減は逆傾向
- $Cov(X, Y) = 0$ のとき, X と Y の増減は無関係



$$Cov(X, Y) > 0$$



$$Cov(X, Y) < 0$$



$$Cov(X, Y) \approx 0$$

共分散 (続き)

183

1. 二つの確率変数 X と Y の和の期待値は, それぞれの期待値の和と等しい:

$$E(X + Y) = E(X) + E(Y)$$

2. しかし X と Y の和の分散は, 一般にはそれぞれの分散の和とは等しくない. 実際,

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

■ 演習: 証明せよ

共分散の使用法の例

184

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

- A社の株価を X , B社の株価を Y とする .

- $Cov(X, Y) > 0$ のとき ,

$$V(X + Y) > V(X) + V(Y)$$

- A , B 両社の株を買うと , 分散が拡大
- 変動リスクが増大し , 資産価値は不安定

- $Cov(X, Y) < 0$ のとき ,

$$V(X + Y) < V(X) + V(Y)$$

- A , B 両社の株を買うと , 分散が縮小
- 変動リスクが抑制され、資産価値は安定

分散・共分散行列

185

- 分散・共分散行列(variance-covariance matrix):

$$\Sigma = E \left[\left\{ \begin{pmatrix} X \\ Y \end{pmatrix} - E \begin{pmatrix} X \\ Y \end{pmatrix} \right\} \left\{ \begin{pmatrix} X \\ Y \end{pmatrix} - E \begin{pmatrix} X \\ Y \end{pmatrix} \right\}^{\top} \right]$$

$$= \begin{pmatrix} V(X) & Cov(X, Y) \\ Cov(Y, X) & V(Y) \end{pmatrix}$$

- 対角成分は分散, 非対角成分は共分散.

- 相関係数(correlation coefficient): 共分散を標準偏差で割った値

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

- 相関係数は $-1 \leq \rho_{XY} \leq 1$ を満たす (証明は宿題) .
- $\rho_{XY} > 0$ のとき, 正の相関があるという .
- $\rho_{XY} < 0$ のとき, 負の相関があるという .
- $\rho_{XY} = 0$ のとき, 無相関(uncorrelated)であるという .
- $|\rho_{XY}| \approx 1$ のとき, X と Y の増減関係はより確定的になる .

- X と Y は互いに独立(independent) :

$$f(x, y) = g(x)h(y)$$

- 2つの確率変数が独立のとき ,

- a. 積の期待値は各々の期待値の積と一致 :

$$E[XY] = E[X]E[Y]$$

- b. 和の積率母関数は各々の積率母関数の積と一致 :

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

- c. 2つの確率変数は無相関 :

$$Cov(X, Y) = 0$$

- 証明は宿題 .

$$M_X(t) = E[e^{tX}]$$

独立性と無相関性

188

- 2つの確率変数が独立ならば無相関である.
- しかし, 逆は一般には正しくない.
- 反例:

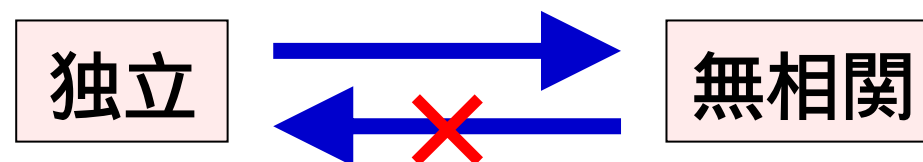
$$X \sim U(-1/2, 1/2), \quad Y = X^2$$

$$E[X] = \int_{-1/2}^{1/2} x dx = 0$$

$$E[XY] = \int_{-1/2}^{1/2} x^3 dx = 0$$

なので $Cov(X, Y) = E[XY] - E[X]E[Y] = 0$

しかし X と Y は明らかに独立ではない.



条件付き確率 (離散型の場合) 189

- 離散型の確率変数 X, Y の条件付き確率分布 (conditional probability distribution):

$Y = y$ が起こったもとで $X = x$ が起こる確率

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$\text{for } P(Y = y) \neq 0$$

- 離散型の確率変数に対する条件付き確率関数 (conditional probability function):

$$p(x|y) = \frac{f(x, y)}{h(y)} \text{ for } h(y) \neq 0$$

条件付き確率(連続型の場合) 190

- 連続型の確率変数 X, Y に対しては, 条件付き確率はどのように定義すればよいか?
- 離散型の定義をそのまま連続型に変えると...

$$P(a \leq X \leq b \mid Y = y) = \frac{P(a \leq X \leq b, Y = y)}{P(Y = y)}$$

- $Y = y$ の一点が起こる確率はゼロなので
 - $P(a \leq X \leq b, Y = y) = 0$
 - $P(Y = y) = 0$

条件付き確率(連続型の場合) 191

- 条件付き確率関数を連続型に変えることにする.
- 条件付き確率密度関数(conditional probability density function):

$$p(x|y) = \frac{f(x, y)}{h(y)} \text{ for } h(y) \neq 0$$

- $p(x|y)$ が確率密度関数になっていることの証明:
 - 非負性: $f(x, y) \geq 0, h(y) > 0$ より $p(x|y) \geq 0$.
 - 正規性: $\int p(x|y)dx = \int \frac{f(x, y)}{h(y)}dx = \frac{h(y)}{h(y)} = 1$

ベイズの定理

192

■ ベイズの定理(Bayes' theorem):

$$p(x|y) = \frac{q(y|x)g(x)}{\int q(y|x)g(x)dx}$$

- $g(x)$: x の**事前確率**(prior probability)
 y を知る前の x の確率
- $p(x|y)$: x の**事後確率**(posterior probability)
 y を知った後の x の確率
- 事前確率 $g(x)$ と条件付き確率 $q(x|y)$ から ,
事後確率 $p(x|y)$ が計算できる

条件付きの期待値と分散

193

- 条件付き確率も確率なので、期待値や分散が定義できる
- 条件付き期待値(conditional expectation):

$$\mu_{X|y} \equiv E(X|y) = \int xg(x|y)dx$$

- 条件付き分散(conditional variance):

$$V(X|y) = \int (x - \mu_{X|y})^2 g(x|y)dx$$

重要な連絡事項

194

- 来週(6月22日)の授業は、海外出張のため
休講です。
- 代わりに(?)宿題をたくさん出しますので、
頑張ってください。

1. $-1 \leq \rho_{XY} \leq 1$ を証明せよ.
2. 二つの確率変数 X と Y が独立のとき, 次式が成り立つことを証明せよ.
 - a. $E[XY] = E[X]E[Y]$
 - b. $M_{X+Y}(t) = M_X(t)M_Y(t)$
 - c. $Cov(X, Y) = 0$

宿題(続き)

196

3. 確率と統計の授業が好きな人, 嫌いな人の確率が

$$P(\text{好})=0.9, \quad P(\text{嫌})=0.1$$

であるとする. また, 確率と統計の授業が好きな人の中で授業中眠たい人, および, 確率と統計の授業が嫌いな人の中で授業中眠たい人の確率が, それぞれ

$$P(\text{眠} \mid \text{好})=0.2, \quad P(\text{眠} \mid \text{嫌})=0.8$$

であるとする.

- $P(\text{好}, \text{眠})$ を求めよ.
- $P(\text{眠})$ を求めよ.
- $P(\text{好} \mid \text{眠})$ を求めよ.
- 「確率と統計の好き嫌い」と「授業中眠たい事」は独立か?

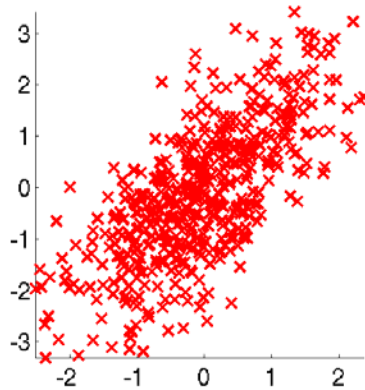
宿題(続き)

197

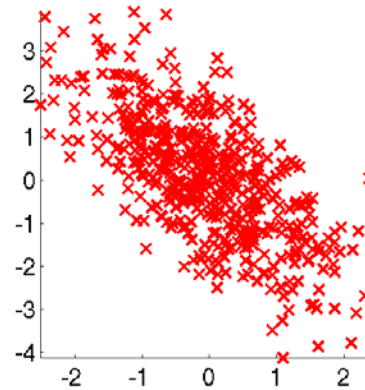
4. Octaveなどを用いて, 次のデータの相関係数を自分で計算してみよ.

<http://sugiyama-www.cs.titech.ac.jp/~sugi/data/Statistics/index-jp.html>

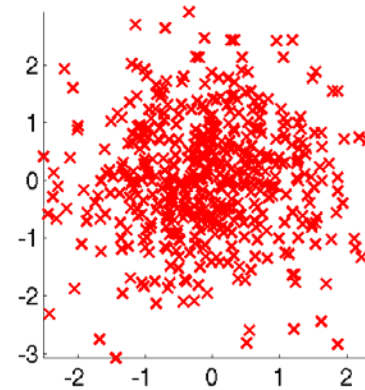
データ1



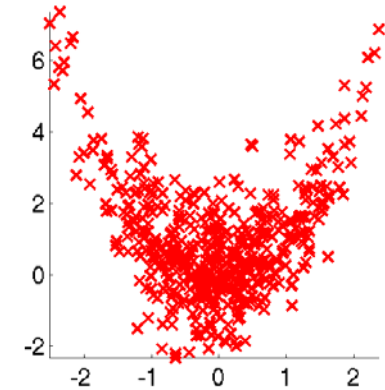
データ2



データ3



データ4



そして, その結果からわかったことを論ぜよ.

5. (余裕のある学生向け): 無相関だけど独立でない
分かりやすい例を考えよ.