確率と統計(O)
「相関と独立性(第7章)」

■担当教員: 杉山 将(計算工学専攻)

■居室: W8E-406

■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/

講義計画(シラバス)

- 確率と統計の基礎
- 確率変数，確率分布
- 積率，積率母関数
- 離散型の確率分布の例
- 連続型の確率分布の例
- 確率不等式， 擬似乱数
- 多次元の確率分布
- 大数の法則，中心極限定理
- 統計的推定，仮説検定

多次元の確率分布

- 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

同時確率

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

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

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

$$f(x, y) \geq 0, \qquad \int\int f(x, y)dxdy = 1$$

周辺確率

- 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．

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

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

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

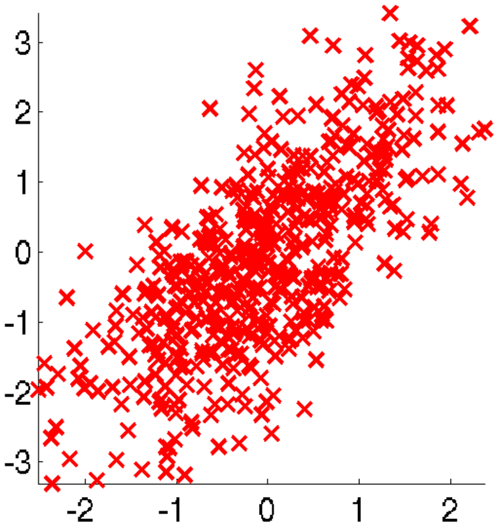
$$g(x) = \int f(x, y)dy \qquad h(y) = \int f(x, y)dx$$

共分散

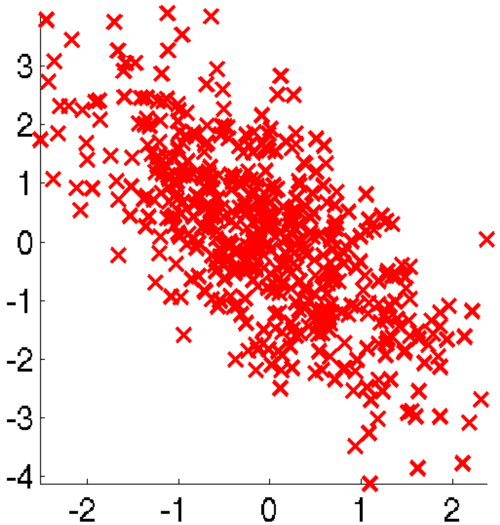
- Xと Yの共分散(covariance):

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

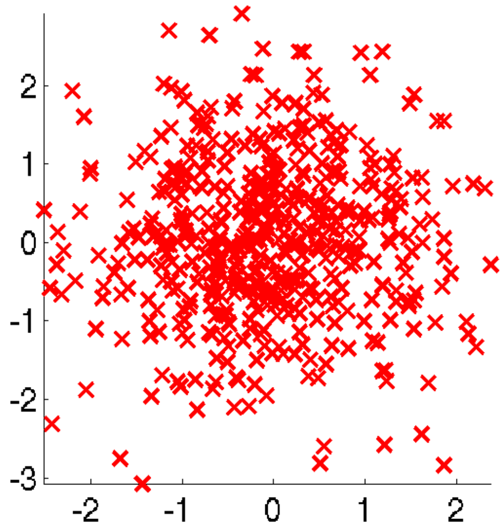
- $Cov(X,Y) > 0$ のとき，X と Y の増減は同傾向
- $Cov(X,Y) < 0$ のとき，X と Y の増減は逆傾向
- $Cov(X,Y) = 0$ のとき，X と Y の増減は無関係



$Cov(X,Y) > 0$



$Cov(X,Y) < 0$



$Cov(X,Y) \approx 0$

共分散(続き)

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

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

- 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

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

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

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

- 対角成分は分散，非対角成分は共分散.

相関

- 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.

独立性

- 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:

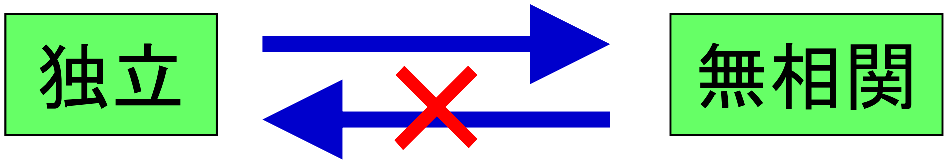
$$Cov(X,Y) = 0$$

- 証明は宿題.

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

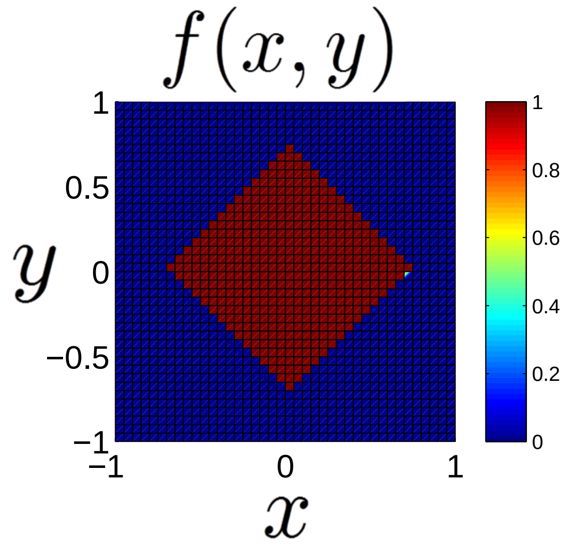
独立性と無相関性

- 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:

$$f(x,y)=\begin{cases}1 & \text{if } |x|+|y|\le\frac{1}{\sqrt{2}}\\ 0 & \text{otherwise}\end{cases}$$



無相関性の証明

$$
\begin{aligned}
Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\
&= E[XY] \qquad \boxed{E[X]=0,\ E[Y]=0} \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} yf(x,y)dydx \\
&= \int_{-\frac{1}{\sqrt{2}}}^{\frac{1}{\sqrt{2}}} x \left(\int_{-\frac{1}{\sqrt{2}}+|x|}^{\frac{1}{\sqrt{2}}-|x|} ydy \right) dx \\
&= -\int_{-\frac{1}{\sqrt{2}}}^{\frac{1}{\sqrt{2}}} \underbrace{\sqrt{2}x|x|dx}_{\text{奇関数}} = 0
\end{aligned}
$$

従属性の証明

- 周辺確率密度関数$g(x)$は,
  - $-\frac{1}{\sqrt{2}} \leq x \leq \frac{1}{\sqrt{2}}$に対して

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\frac{1}{\sqrt{2}}+|x|}^{\frac{1}{\sqrt{2}}-|x|} dy = \sqrt{2} - 2|x|$$

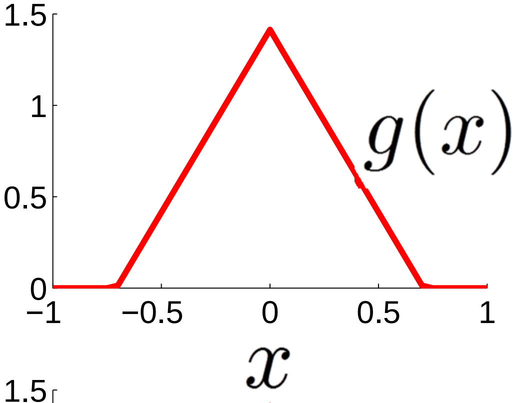
  - それ以外の場合は

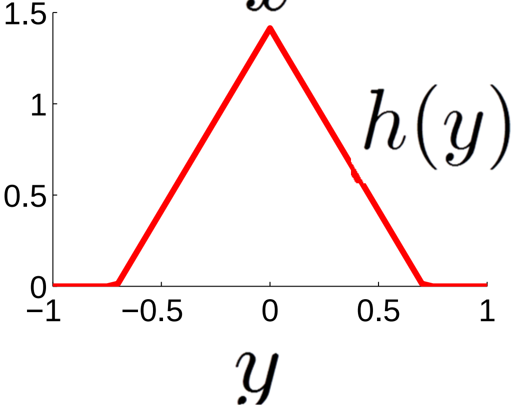
$$g(x) = 0$$

よって, $g(x) = \max(0, \sqrt{2} - 2|x|)$

同様に,

$$h(y) = \max(0, \sqrt{2} - 2|y|)$$





従属性の証明

■ 従属性の証明:

- 同時確率密度関数は

$$f(x,y) = \begin{cases} 1 & \text{if } |x|+|y| \le \frac{1}{\sqrt{2}} \\ 0 & \text{otherwise} \end{cases}$$

- 周辺確率密度関数の積は

$$g(x)h(y) = \begin{cases} (\sqrt{2}-2|x|)(\sqrt{2}-2|y|) & \text{if } -\frac{1}{\sqrt{2}} \le x, y \le \frac{1}{\sqrt{2}} \\ 0 & \text{otherwise} \end{cases}$$

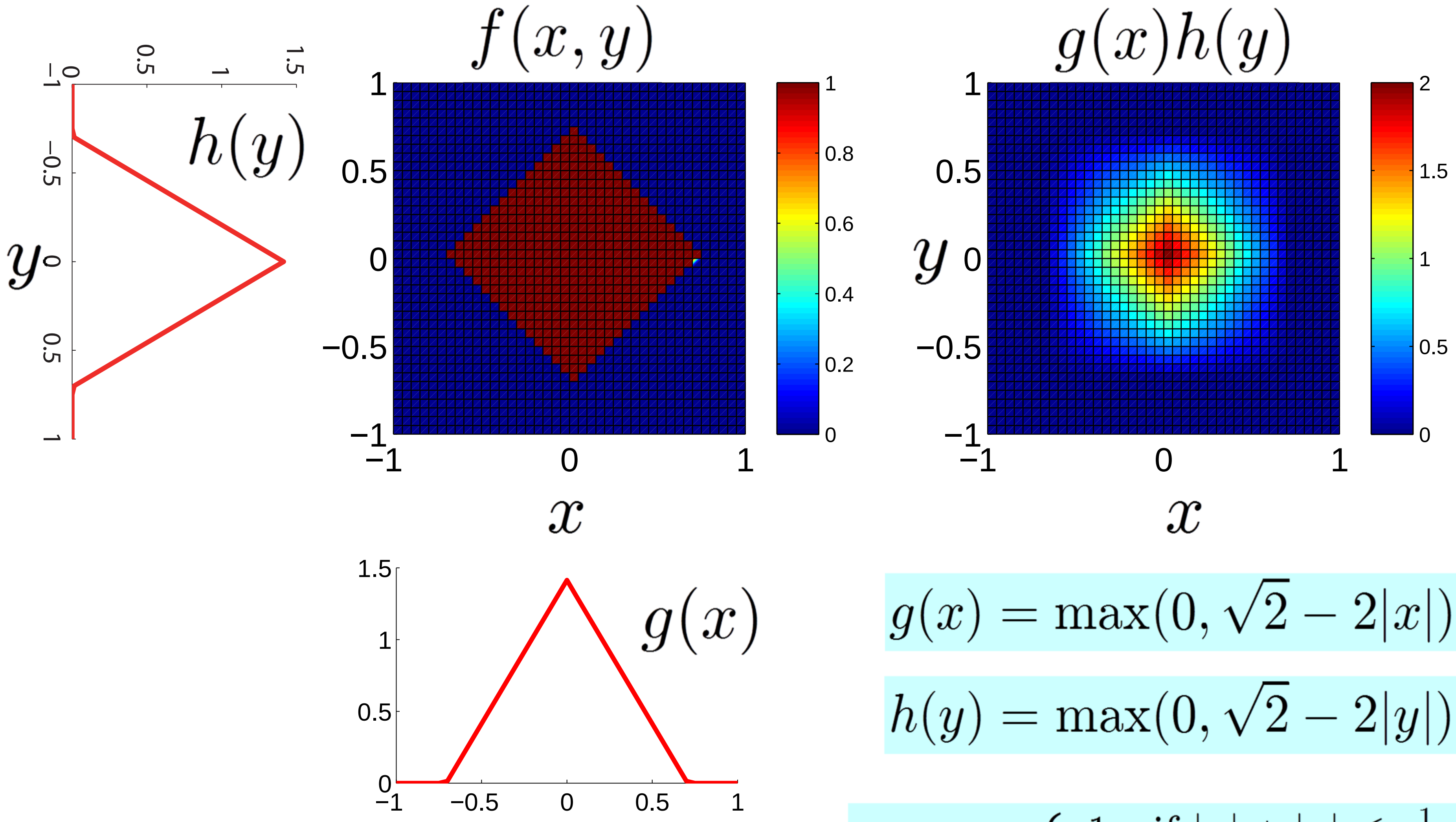
$$g(x) = \max(0, \sqrt{2}-2|x|)$$

$$h(y) = \max(0, \sqrt{2}-2|y|)$$

- 従って,

$$g(x)h(y) \neq f(x,y)$$

従属性の証明



$$g(x) = \max(0, \sqrt{2} - 2|x|)$$

$$h(y) = \max(0, \sqrt{2} - 2|y|)$$

$$f(x,y) = \begin{cases} 1 & \text{if } |x| + |y| \leq \frac{1}{\sqrt{2}} \\ 0 & \text{otherwise} \end{cases}$$

まとめ

- 同時確率
- 周辺確率
- 共分散
- 相関
- 独立性と無相関性

宿題

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. $Cov(X, Y) = 0$

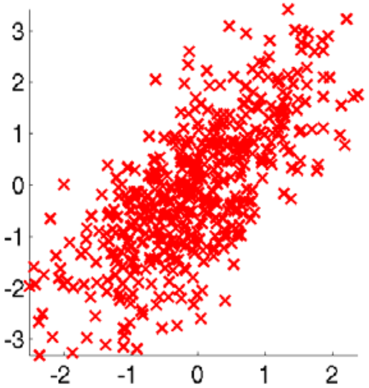
$$M_X(t) = E[e^{tX}]$$

宿題(続き)

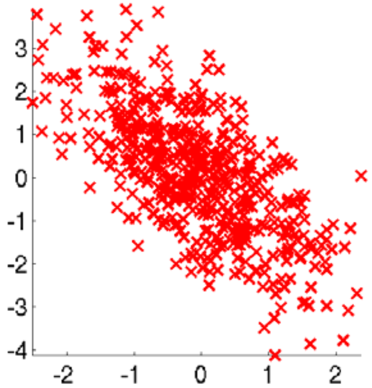
3. Octaveなどを用いて，次のデータの相関係数を自分で計算してみよ.

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

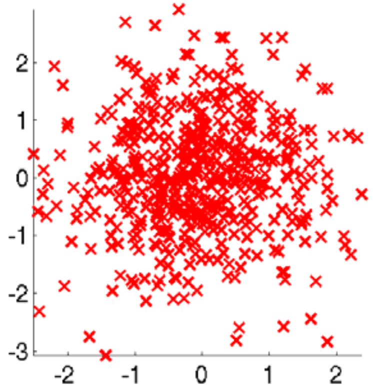
データ1



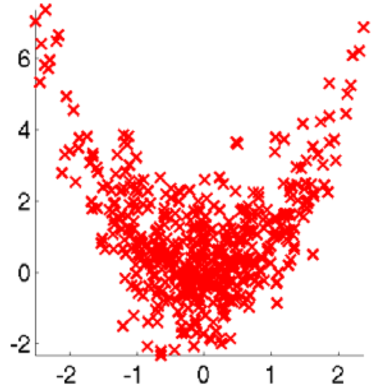
データ2



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.