データ解析 (第4回)

静岡大学システム工学科
安藤 和敏

(4) 2次元分布

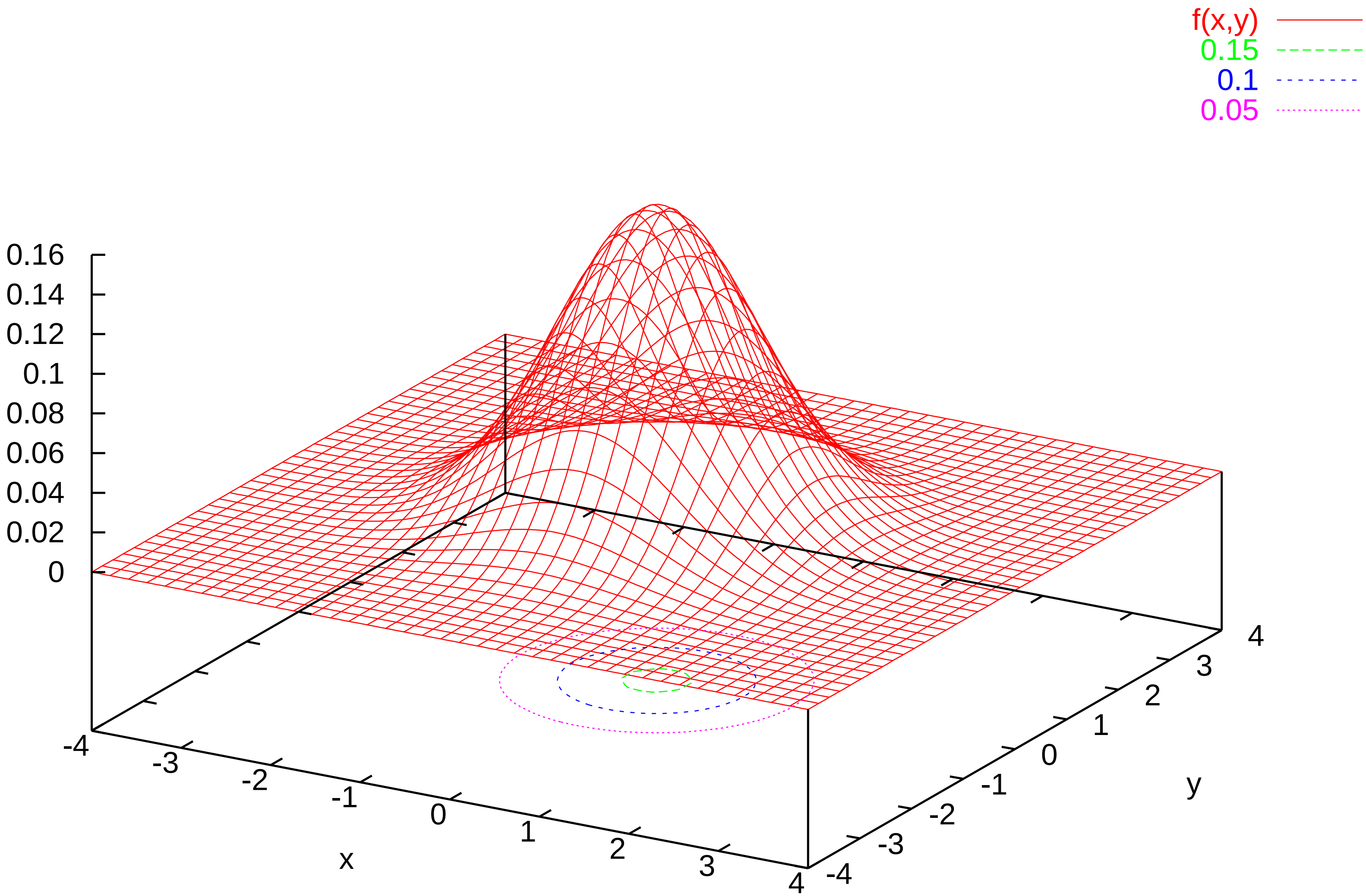
同時確率密度関数

以下の 2 つの条件を満たす $f(x,y)$ を (x,y) の同時確率密度関数と呼ぶ.

$$f(x,y) \geq 0 \quad (-\infty < x, y < \infty) \tag{2.36}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = 1 \tag{2.37}$$

同時確率密度関数のグラフの例



周辺密度関数

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y \qquad (2.38)$$

を x の周辺確率密度関数と呼び,

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}x \qquad (2.38)$$

を y の周辺確率密度関数と呼ぶ.

$F(x,y)$ を x と y の任意の関数とする. このとき, $F(x,y)$ の期待値は,

$$E\{F(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$$

と定義される.

特に, $F(x,y)=x$ の場合は,

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x \\ &= \int_{-\infty}^{\infty} x g(x)\mathrm{d}x \quad \text{(2.41).} \end{aligned}$$

また, $F(x,y)=ax+by$ の場合は,

$$\begin{aligned}
E(ax+by) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)f(x,y)\mathrm{d}x\mathrm{dy} \\
&= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)\mathrm{d}x\mathrm{dy} \\
&\quad +b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)\mathrm{d}x\mathrm{dy} \\
&= aE(x)+bE(y) \quad (\text{(2.41) より})
\end{aligned}$$

共分散の定義

$$C(x,y) = E\{(x-E(x))(y-E(y))\} \quad (2.43)$$

母相関係数の定義

$$\rho_{xy} = \frac{C(x,y)}{V(x)V(y)} \quad (2.44)$$

共分散の性質

$$C(x, y) = E(xy) - E(x)E(y) \tag{2.45}$$

$$V(ax + by) = a^2V(x) + b^2V(y) + 2abC(x, y) \tag{2.46}$$

(2.45) の証明

$$
\begin{aligned}
&C(x, y) \\
&= E\{(x - E(x))(y - E(y))\} \quad (\text{定義 (2.43) より} \\
&= E\{xy - xE(y) - yE(x) + E(x)E(y)\} \\
&= E(xy) - E(x)E(y) - E(y)E(x) + E(x)E(y \\
&= E(xy) - E(x)E(y).
\end{aligned}
$$

(2.46) の証明

$$
\begin{aligned}
&V(ax+by) \\
&\quad = \ E\{(ax+by-E(ax+by))^2\} \quad \text{(定義 (2.25) よ} \\
&\quad = \ E\{(a(x-E(x))+b(y-E(y)))^2\} \quad \text{((2.42) よ} \\
&\quad = \ E\{a^2(x-E(x))^2+b^2(y-E(y))^2 \\
&\qquad\quad +2ab(x-E(x))(y-E(y))\} \\
&\quad = \ a^2E\{(x-E(x))^2\}+b^2E\{(y-E(y))^2\} \\
&\qquad\quad +2abE\{(x-E(x))(y-E(y))\} \quad \text{((2.42) よ} \\
&\quad = \ a^2V(x)+b^2V(y)+2abC(x,y). \quad \text{((2.25),(2.4}
\end{aligned}
$$

確率変数の独立性

同時確率密度関数 $f(x,y)$ と周辺密度関数 $g(x), h(y)$ のあいだに,

$$f(x,y) = g(x)h(y) \tag{2.47}$$

という関係があるとき, x と y は独立であるという.

x と y が独立な場合の性質

$$E(xy) = E(x)E(y) \tag{2.48}$$

$$C(x, y) = 0 \tag{2.49}$$

$$\rho_{xy} = 0 \tag{2.50}$$

$$V(ax + by) = a^2V(x) + b^2V(y) \tag{2.51}$$

(2.48) が成り立つことさえ分かれば，(2.49)〜(2.51) は, (2.44)〜(2.46) からすぐにわかる.

(2.48)の証明

$$
\begin{aligned}
E(xy) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xg(x)yh(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty} xg(x)\mathrm{d}x \cdot \int_{-\infty}^{\infty} yh(y)\mathrm{d}y \\
&= E(x)\cdot E(y).
\end{aligned}
$$