2 多変量分布

2.1 結合分布

まず，2 変量の場合を考える．確率変数を X と Y とし，その**累積分布関数**（cumulative distribution function: cdf）を

$$F(x,y) = \Pr(X \le x, Y \le y) \quad x, y \in \mathbb{R}$$

とする．この講義では，$F(x,y)$ が絶対連続の場合を考える．つまり，ほとんどいたるところで

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y) \tag{1}$$

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)dudv$$

が成立する．このとき，(1) の $f(x,y)$ を**確率密度関数**（probability density function: pdf）と呼ぶ．

確率密度関数の性質

- $f(x,y) \ge 0$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)dudv = 1$

例えば，確率変数 X, Y が長方形の中に含まれる確率は，

$$\begin{aligned}
&\Pr(x \le X \le x+\Delta x, y \le Y \le y+\Delta y)\\
&\quad = F(x+\Delta x, y+\Delta y) - F(x+\Delta x, y) - F(x, y+\Delta y) + F(x,y)\\
&\quad = \boxed{\text{読者の演習}}\\
&\quad = \int_{y}^{y+\Delta y}\int_{x}^{x+\Delta x} f(u,v)dudv
\end{aligned}$$

のように多重積分で表される．ただし，$\Delta x > 0, \Delta y > 0$ である．また，確率変数 X, Y が可測集合 E に含まれる確率は，以下で与えられる：

$$\Pr((X,Y) \in E) = \int\int_{E} f(x,y)dxdy.$$

次に，p 変量の場合を考える．確率変数を $X_1, \dots, X_p$ とし，その cdf を

$$F(x_1, \dots, x_p) = \Pr(X_1 \leq x_1, \dots, X_p \leq x_p) \quad x_1, \dots, x_p \in \mathbb{R}$$

とする．もし $F(x_1, \dots, x_p)$ が絶対連続ならば，ほとんどいたるところで

$$\frac{\partial^p F(x_1, \dots, x_p)}{\partial x_1 \cdots \partial x_p} = f(x_1, \dots, x_p) \tag{2}$$

が成り立ち，

$$F(x_1, \dots, x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, \dots, u_p) du_1 \cdots du_p$$

である．このとき，(2) の $f(x_1, \dots, x_p)$ は pdf である．

2.2 周辺分布

まず，2変量の場合を考える．確率変数を X と Y とし，その cdf を $F(x, y)$ とする．このとき，

$$\begin{aligned} \Pr(X \leq x) &= \Pr(X \leq x, Y \leq \infty) \\ &= F(x, \infty) \end{aligned}$$

である．この $F(x, \infty)$ を $F(x)$ と記すと，明らかに

$$F(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v) dv du \tag{3}$$

が成り立ち，$F(x)$ を X の**周辺累積分布関数**（marginal cdf）という．また，

$$\int_{-\infty}^{\infty} f(u, v) dv = f_X(u) \tag{4}$$

を**周辺確率密度関数**（marginal pdf）という．さらに，(4) より (3) は次のようにも表せる：

$$F(x) = \int_{-\infty}^{x} f_X(u) du.$$

同様にして，Y の周辺 cdf と周辺 pdf をそれぞれ $G(y)$ と $f_Y(y)$ とする．

次に，p 変量の場合を考える．確率変数を $X_1, \ldots, X_p$ とし，その cdf を $F(x_1, \ldots, x_p)$ とする．$X_1, \ldots, X_r$（$r < p$）の周辺 cdf は，

$$\begin{aligned}
&\Pr(X_1 \le x_1, \ldots, X_r \le x_r) \\
&\quad = \Pr(X_1 \le x_1, \ldots, X_r \le x_r, X_{r+1} \le \infty, \ldots, X_p \le \infty) \\
&\quad = F(x_1, \ldots, x_r, \infty, \ldots, \infty)
\end{aligned}$$

であり，$X_1, \ldots, X_r$ の周辺 pdf は，

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_p) du_{r+1} \cdots du_p$$

である．$X_1, \ldots, X_p$ の任意の部分集合に対する周辺 cdf や周辺 pdf も同様に与えられる．

2.3 統計的独立

まず，2 変量の場合を考える．確率変数 X, Y の cdf を $F(x, y)$，X と Y の周辺 cdf をそれぞれ $F(x)$ と $G(y)$ とするとき，

$$F(x, y) = F(x)G(y) \quad x, y \in \mathbb{R} \tag{5}$$

が成り立つならば，X と Y は**互いに独立**（mutually independent）であるという．また，(5) は次のようにも表せる：

$$f(x, y) = f_X(x)f_Y(y) \quad x, y \in \mathbb{R} \tag{6}$$

(5) と (6) が同値であることの証明

(5) ⇒ (6) について

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{dF(x)}{dx}\frac{dG(y)}{dy} = f_X(x)f_Y(y)$$

(6) ⇒ (5) について

$$F(x, y) = \boxed{\text{読者の演習}} = F(x)G(y)$$

次に，p 変量の場合を考える．確率変数 $X_1,\dots,X_p$ の cdf を $F(x_1,\dots,x_p)$，$F_i(x_i)$ を X_i（$i=1,\dots,p$）の周辺 cdf とするとき，

$$F(x_1,\dots,x_p)=F_1(x_1)\cdots F_p(x_p)\quad x_1,\dots,x_p\in\mathbb{R}$$

が成り立つならば，$X_1,\dots,X_p$ は互いに独立であるという．また，

$$F(x_1,\dots,x_p)=F(x_1,\dots,x_r,\infty,\dots,\infty)\cdot F(\infty,\dots,\infty,x_{r+1},\dots,x_p)$$

が成り立つならば，確率変数 $X_1,\dots,X_r$ と確率変数 $X_{r+1},\dots,X_p$ は独立であるという．

2.4 条件付き分布

まず，2 変量の場合を考える．確率変数 X,Y の cdf を $F(x,y)$ とする．$Y=y$ を与えた X の**条件付き確率密度関数**（conditional pdf）は，

$$\frac{f(x,y)}{\displaystyle\int_{-\infty}^{\infty}f(u,y)du}=\frac{f(x,y)}{f_Y(y)} \tag{7}$$

で与えられる．(7) の $f(x,y)/f_Y(y)$ を $f(x|y)$ と記すこともある．また，$Y=y$ を与えたとき X が x_1 以上かつ x_2 以下である条件付き確率は，

$$\Pr(x_1\le X\le x_2|Y=y)=\int_{x_1}^{x_2}f(u|y)du$$

で与えられる．

条件付き確率密度関数の性質

- $f(x|y)\ge 0$
- $\displaystyle\int_{-\infty}^{\infty}f(u|y)du=\frac{1}{f_Y(y)}\int_{-\infty}^{\infty}f(u,y)du=\frac{f_Y(y)}{f_Y(y)}=1$

次に，p 変量の場合を考える．確率変数 $X_1,\dots,X_p$ の cdf を $F(x_1,\dots,x_p)$ とする．$X_{r+1}=x_{r+1},\dots,X_p=x_p$ を与えた $X_1,\dots,X_r$ の条件付き確率密度関数は，

$$\frac{f(x_1,\dots,x_p)}{\displaystyle\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(u_1,\dots,u_r,x_{r+1},\dots,x_p)du_1\dots du_r}$$

で与えられる．

2.5 演習問題

問 1 確率変数 X, Y の pdf が

$$f(x,y) = \begin{cases} e^{-(x+y)} & (x>0,\ y>0) \\ 0 & (else) \end{cases}$$

であるとき，確率変数 X, Y の cdf は（ア）である．文中の（ア）に当てはまるものとして，次の ① ～ ④ のうちから適切なものを一つ選べ． $\boxed{\textbf{1}}$

① 1

② $1 - e^{-x}$

③ $1 - e^{-x} - e^{-y}$

④ $1 - e^{-x} - e^{-y} + e^{-(x+y)}$

問 2 確率変数 X, Y の pdf が

$$f(x,y) = \begin{cases} \dfrac{21}{4}x^2 y & (x^2 \le y \le 1) \\ 0 & (else) \end{cases}$$

であるとき，X の周辺 pdf を求める．X の値域は $-1 \le x \le 1$ であるから，$|x| > 1$ のとき $f_X(x) = 0$ である．また，

$$f_X(x) = \int_{(イ)}^{1} \frac{21}{4}x^2 y dy = \frac{21}{8}x^2(1-x^4) \quad (|x| \le 1).$$

さらに，$X = 1/2$ を与えたとき Y が $3/4$ 以上となる条件付き確率は（ウ）である．

1. 文中の（イ）に当てはまるものとして，次の ① ～ ④ のうちから適切なものを一つ選べ． $\boxed{\textbf{2}}$

   ① -1　② 0　③ x　④ x^2

2. 文中の（ウ）に当てはまるものとして，次の ① ～ ④ のうちから適切なものを一つ選べ． $\boxed{\textbf{3}}$

   ① $1/15$　② $4/15$　③ $7/15$　④ $11/15$