2020.5.8(金)

情報統計力学 学習４

5月12日 (火) 午後１時ー２時半

もう少し、数学的な話をします。概念を理解してほしいからです。その後で、大数の法則、中心極限定理、大偏差原理を解説します。情報エントロピー、物理で定義されるエントロピーとの関係などは、新型コロナウィルス禍が収束後、教室でお話できると期待しています。

確率変数

まず、厳密な定義を記す。

定義
空間 Ω 上の実数値関数 X が次の条件を満たすとき、X は**確率変数**であるという。

任意の実数 a に対して、$\{\omega \in \Omega; X(\omega) \leq a\}$ は事象である。

通常、確率変数は大文字 (X) で書き、それがとる実数値を小文字 (x) で書く。

なんだか、取ってつけたような定義と思うかもしれないが、数学的に厳密に定義するとこうなるのである。こう定義すると、X がある値や、ある区間にある確率が決まる。つまり、確率変数 X が取りうる値 x と、その x をとる確率が同時に決まるのである。

確率は、事象 $A \in \mathcal{F}$ について定義されていたことを思いだそう。
今後、省略して、事象 $\{\omega \in \Omega; X(\omega) \leq a\}$ の代わりに、$\{X \leq a\}$ を用いる。他も同様。

確率変数の分布を次の定義で与える。

定義
確率変数 X によって実数空間（実数の全体)$\mathbb{R}$ 上に誘導される（確率）分布 P_X を確率変数 X の**分布** (distribution) または**法則** (law) という。すなわち、

$$P_X(A) = P(X \in A) = P(\{\omega \in \Omega, X(\omega) \in A\}),\ A \subset \mathbb{R}$$

離散型確率変数

確率変数 X が可算個の値のみを取るとき、X を**離散型確率変数**という。この場合、

$$P(X = x_i) = p_i\ (i = 1, 2, \cdots)$$

$$p_i \geq 0,\ \sum_{i=1}^{\infty} p_i = 1$$

となる。これを**離散型確率分布**という。特に、

$$p(x) := P(X = x), (x = x_1, x_2, \cdots)$$

を確率変数 X の**確率関数**という。

次に確率変数が連続の場合を考える。

連続型確率変数

定義

確率変数の取る値が連続量となる場合に、X のことを**連続型確率変数**という。この場合、非負関数 $f(x)$ が存在して、任意の**区間** $A \subset \mathbb{R}$ に対して、

$$P(X \in A) = \int_A f(x)dx$$

なる関係が成り立つ。この確率分布を**連続型確率分布**といい、$f(x)$ を確率変数 X の**確率密度関数**あるいは、単に **密度関数**という。

全確率は 1 なので、

$$P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x)dx = 1$$

となっている。また、X の値がある区間 $[a, b]$ に入っている確率は、定義から

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

となる。

例 正規密度関数
連続型確率変数 X が密度関数として次の正規密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}], \ -\infty < x < \infty \tag{1}$$

をもつ場合がある。対応する確率分布は標準正規分布と呼ばれ、記号で $N(0, 1)$ と書かれる。

確率変数 X がある分布 B に従っているとき、記号で $X \sim B$ と書く。従って、標準正規分布に従っているときには、

$$X \sim N(0, 1)$$

と書く。

分布関数

確率変数 X に対して、X の **(確率) 分布関数** F を次のように定義する。

$$F(x) \equiv F_X(x) = P(X \leq x) = P(\{\omega \in \Omega; X(\omega) \leq x\}), \ (-\infty < x < \infty). \tag{2}$$

F の添字 X は、F が「確率変数 X の分布関数である」ことを強調するために書かれており、文脈からそれが明らかな場合は省略してよい。統計学では、しばしば**累積分布関数**とも呼ばれる。

確率変数 X が離散型の場合。

$$F(x) = \sum_{i: x_i < x} p_i.$$

確率変数 X が連続型の場合。

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

これまでは、確率変数が１つの場合であったが、多数の場合への拡張は容易である。ここでは、２変数の場合を解説する。

２次元確率ベクトルの場合

同じ確率空間 $(\Omega, \mathcal{F}, P)$ 上で定義された２つの確率変数 X, Y を考える。このとき、点 $(X,Y) \equiv (X(\omega), Y(\omega))$ が２次元平面 $\mathbb{R}^2$ 上の任意の区間 $I = \{(x,y) \in \mathbb{R}^2; a < x \le b, c < y \le d\}$ に入るような $\omega \in \Omega$ の全体は $\mathcal{F}$ に属する。その理由は以下の通りである。

$$\{\omega; (X(\omega), Y(\omega)) \in I\} = \{\omega; a < X(\omega) \le b\} \cap \{\omega; c < Y(\omega)) \le d\}$$

となる。一方、

$$\{\omega; a < X(\omega) \le b\} = \{\omega; X(\omega) \le b\} \cap \{\omega; a < X(\omega)\},$$

$$\{\omega; a < X(\omega)\} = \{\omega; X(\omega) \le a\}^c$$

である。確率変数の定義より $\{\omega; X(\omega) \le b\} \in \mathcal{F}$ である。また、$\{\omega; a < X(\omega)\} = \{\omega; X(\omega) \le a\}^c \in \mathcal{F}$ であるが、$\mathcal{F}$ の要素の補修合は、$\mathcal{F}$ の元であるから、$\{\omega; a < X(\omega)\}$ は $\mathcal{F}$ の元である。また、$\mathcal{F}$ の元の共通部分も $\mathcal{F}$ の元であるから、$\{\omega; a < X(\omega) \le b\} \in \mathcal{F}$ となる。同様にして、$\{\omega; c < Y(\omega)) \le d\} \in \mathcal{F}$ である。よって、この２つの共通部分である $\{\omega; (X(\omega), Y(\omega)) \in I\}$ も $\mathcal{F}$ の元、すなわち、事象である。従ってその確率を計算できる。
組 (X,Y) を $(\Omega, \mathcal{F}, P)$ 上の**２次元確率ベクトル**という。任意の点 $(x,y) \in \mathbb{R}^2$ に対して

$$F(x,y) = F_{XY}(x,y) = P(X \le x, Y \le y) = P(\{\omega \in \Omega; X(\omega) \le x, Y(\omega) \le y\})$$

で定義される $F_{X,Y}$ を (X,Y) の**２次元分布関数**、あるいは X と Y の**同時分布関数**という。
同時分布関数の基本的な性質を以下に示す。

1. $x_1 < x_2, y_1 < y_2$ ならば、$F_{XY}(x_1, y_1) \le F_{XY}(x_2, y_2)$.
2. F_X を X の分布関数、F_Y を Y の分布関数とするとき、
   任意の $y \in \mathbb{R}$ に対して、$F_{XY}(\infty, y) = F_Y(y)$
   任意の $x \in \mathbb{R}$ に対して、$F_{XY}(x, \infty) = F_X(x)$
3. $F_{XY}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$
4. $F_{XY}(\infty, \infty) = 1$
5. $P(a < X \le b, c < Y \le d) = F_{XY}(b,d) - F_{X,Y}(b,c) - F_{XY}(a,d) + F_{XY}(a,c)$

$F_X(x) = F_{XY}(x, \infty)$ を X **の周辺分布関数**、$F_Y(y) = F_{XY}(\infty, y)$ を Y **の周辺分布関数**という。

上の性質の証明を試みて下さい。

X, Y が離散型確率変数の場合は、それらの値域をそれぞれ $\{x_i; i = 1, 2, \cdots\}, \{y_j; j = 1, 2, \cdots\}$ とすれば、$(X(\omega), Y(\omega)) = (x_i, y_j)$ に対する確率

$$p_{ij} \quad = \quad P(X = x_i, Y = y_j),\ (i = 1, 2, \cdots;\ j = 1, 2, \cdots) \tag{3}$$

が決まる。$x_1 < x_2 < \cdots, y_1 < y_2 < \cdots$ とする。このとき、

$$p_{ij} \geq 0,\ \sum_{i,j=1}^{\infty} p_{ij} = 1$$

が成り立つ。$p_{i.j}$ は X, Y の**同時確率分布**とよばれる。同時分布関数は

$$F_{XY}(x, y) = \sum_{x_i \leq x, y_j \leq y} p_{ij} \tag{4}$$

となる。確率関数は、$p(x, y) = P(X = x, Y = y)$ と定義され、次式が成立する。

$$p(x, y) \geq 0,\ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p(x_i, y_j) = 1,\ F_{XY}(x, y) = \sum_{x_i \leq x}\sum_{y_j \leq y} p(x_i, x_i). \tag{5}$$

X, Y の周辺分布 $p_{i,\cdot}, p_{\cdot,j}$ は、

$$p_{i,\cdot} = \sum_{j=1}^{\infty} p_{ij} = P(X = x_i), \tag{6}$$

$$p_{\cdot,j} = \sum_{i=1}^{\infty} p_{ij} = P(Y = y_j) \tag{7}$$

となる。

X, Y が連続型確率変数の場合。

$$P(a < X \leq b, c < Y \leq d) = \int_a^b dx \int_c^d dy f(x, y) \tag{8}$$

となるような関数 $f(x, y)$ が存在するとき、それを**同時密度関数**という。

$$f(x, y) \geq 0,\ \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy f(x, y) = 1$$

であり、同時分布関数との関係は、

$$F(x, y) = \int_{-\infty}^{x} du \int_{-\infty}^{y} dw f(u, w)$$

となる。X, Y の周辺分布の密度関数はそれぞれ

$$f(x, \cdot) = \int_{-\infty}^{\infty} dw f(x, w),\ f(\cdot, y) = \int_{-\infty}^{\infty} du f(u, y)$$

となる。周辺分布関数について、次式が成り立つ。

$$F_{XY}(x, \infty) = \int_{-\infty}^{x} du f(u, \cdot),\ F_{XY}(\infty, y) = \int_{-\infty}^{y} dw f(\cdot, w)$$

独立性

２つの確率変数 X, Y について、$a < b, c < d$ を満たす任意の実数 a, b, c, d に対して、

$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b)P(c < Y \leq d) \tag{9}$$

が成り立つとき、X と Y は**互いに独立である**という。

定理
$Z=(X,Y)$ を 2 次元確率変数とする。Z が離散型のとき、Z,X,Y の確率関数をそれぞれ p,p_X,p_Y とすると、確率変数 X と Y が独立であるための必要十分条件は

$$p(x,y)=p_X(x)p_Y(y)$$

となることである。

定理
$Z=(X,Y)$ を 2 次元確率変数とする。Z が連続型のとき、Z,X,Y の確率密度関数をそれぞれ $f(x,y),f_X(x),f_Y(y)$ とすｒ．確率変数 X と Y が独立であるための必要十分条件は

$$f(x,y)=f_X(x)f_Y(y)$$

となることである。