数理統計学まとめ（その5）：第 3 章 確率変数と確率分布

正規分布表の読み方

Z が標準正規分布に従うとき $P(Z \geq a) = 0.08$ となる a を標準正規分布表から求めてみる．

$$P(Z \geq a) = P(0 \leq Z < \infty) - P(0 < Z < a) = 0.5 - P(0 < Z < a)$$

だから $P(0 < X < a) = 0.42$ を満たす a を正規分布表で探す．$P(0 \leq Z \leq 1.40) = 0.4192$ で $P(0 \leq Z \leq 1.41) = 0.4207$ だから a は 1.40 と 1.41 の間．

$$f(a) = P(0 \leq Z \leq a)$$

が $a = 1.40$ と 1.41 の間で一次式になっているとして計算する．

$$f(a) = c(a - 1.40) + 0.4192 \qquad (f(1.40) = 0.4192 \text{ を満たすので })$$

とかけるとして，c を求める．表から $f(1.41) = 0.4207$ だったので，

$$0.4207 = c \times 0.01 + 0.4192$$

これを解くと $c = 0.15$ なので，$f(a) = 0.15 \times (a - 1.40) + 0.4192$ となり，$f(a) = 0.42$ となる a はこれから

$$a - 1.40 = (0.42 - 0.4192)/0.15 = 0.005(\text{小数第 4 位 四捨五入}) \quad \therefore a = 1.405$$

と計算する．

5　2 次元の確率変数

離散確率変数のとき

二つのデータの組 $\{(x_j, y_j)\}_{j=1}^{n}$ を考えたように，二つの確率変数 X, Y を組 (X, Y) として考えることも多い．X, Y のそれぞれが指定された値 x, y をとる確率

$$P(X = x, Y = y)$$

を X, Y がとりうる可能な値の組 x, y すべてについて与えたものを X, Y の**同時分布**という．

例 5.1 （教科書例題 3.15）二つのサイコロがある．X, Y をそれぞれ第 1 および第 2 のサイコロの出る目の数とするとき，次の 2 次元の確率変数の分布を考えよ．

(1) (X, Y)

(2) $A = X + Y$, $D = |X - Y|$

(1) については X, Y それぞれ 1 から 6 の値をとり，どの組み合わせも同じ確率 1/36 で起こるので，

$$P\left((X, Y) = (i, j)\right) = P(X = i, Y = j) = \frac{1}{36}$$

が $i, j = 1, 2, \ldots, 6$ に対して成り立っている．これで同時分布が決まっている．

(2) の A, D の同時分布は次のような表になる．

D＼A	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$
1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0
2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0
3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0
4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0
5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0

この表の各行で横に足すと，D の分布 (D の**周辺分布**という）各列で縦に足すと A の周辺分布が得られる．一般に（離散）確率変数 X, Y の同時分布が与えられているとき X, Y それぞれの周辺分布は次式で与えられる．

$$P(X = x_i) = \sum_{j=1}^{m} P(X = x_i, Y = y_j), \quad P(Y = y_j) = \sum_{i=1}^{n} P(X = x_j, Y = y_j)$$

上の例の A と D の周辺分布については次のようになる．

A	2	3	4	5	6	7	8	9	10	11	12
確率	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

D	0	1	2	3	4	5
確率	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

定義 5.1 二つの (離散) 確率変数 X, Y が**独立**とは (X, Y) の取りうるすべての値の組 (x_i, y_j) について

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

が成り立つ事を言う．X, Y が独立でないとき **従属**であるという.

上の例で X, Y は明らかに独立だが A, D は従属である．なぜなら

$$P(A = 3, D = 5) = 0 \text{ だが } P(A = 3)P(D = 5) = \frac{2}{36} \times \frac{2}{36} \neq 0$$

だから.

連続型確率変数のとき

二つの確率変数 X, Y がともに連続型のとき，その同時分布は $a \leq b, c \leq d$ に対し，$P(a < X \leq b, c < Y \leq d)$ で与えられる．特に

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

となるとき $f(x, y)$ を (X, Y) の **同時確率密度関数** という．この積分は x と y について別々に行う（**重積分**と言う．微積で後期に習う）．これを簡単に

$$\int_a^b \int_c^d f(x, y) dx dy$$

と書くことも多い.

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

は X の **周辺分布密度関数**といい,

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

は Y の **周辺分布密度関数**と言う．

同時確率密度関数 $f(x,y)$ が

$$f(x,y) = f_1(x)f_2(y)$$

と周辺分布密度関数の積になるとき X,Y は**独立**であると言い，そうでないとき**従属**であるという． 2 次元の確率変数の関数 $g(X,Y)$ の**平均**は

$$E(g(X,Y)) = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i,y_j)P(X=x_i,Y=y_j) \qquad \text{(離散型)}$$

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \qquad \text{(連続型)}$$

と定義する．$E(X)=\mu_X$，$E(Y)=\mu_Y$ と書く．連続型のときは

$$E(X) = \int_{-\infty}^{\infty} xf_1(x)\,dx, \quad E(Y) = \int_{-\infty}^{\infty} yf_2(y)\,dy$$

となる．このとき，$g(X,Y)=(X-\mu_X)(Y-\mu_Y)$ の平均を X,Y の**共分散**と呼び，$\mathrm{Cov}(X,Y)$ で表し，結果を $\sigma_{X,Y}$ と書く．

$$\mathrm{Cov}(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i-\mu_X)(y_j-\mu_Y)P(X=x_i,Y=y_j) \qquad \text{(離散型)}$$

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy \qquad \text{(連続型)}$$

共分散 $\mathrm{Cov}(X,Y)$ を X,Y の標準偏差 σ_X,σ_Y でわったもの

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y}$$

を X,Y の**相関係数**と呼ぶ．

定理 5.1 （教科書 p.73 定理 3.5）X,Y が独立ならば

$$\mathrm{Cov}(X,Y) = 0$$