2004 年 6 月 22 日

総合科目　数理・自然 6 「統計学入門」講義レジュメ　(今野)

離散型確率変数の同時確率分布

X と Y を離散型確率変数とする．

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

を 2 次元確率変数 (X, Y) の同時確率分布という．

性質

1.
$$f(x, y) \geq 0, \qquad \sum_x \sum_y f(x, y) = 1$$

2. 2 次元空間の中にある事象とは (x, y) の集まった部分集合である．事象 A に対して
$$P\{(X, Y) \in A\} = \sum \sum_{(x, y) \in A} f(x, y)$$

離散型の周辺分布

X と Y の周辺確率分布を

$$g(x) = \sum_y f(x, y), \qquad h(y) = \sum_x f(x, y)$$

で定義する．

連続型確率変数の同時確率分布

X と Y を連続型確率変数とする． $f(x, y)$ は 2 次元の同時確率密度関数とする．

性質

1.
$$f(x, y) \geq 0, \qquad \int \int_S f(x, y)\, dx\, dy = 1$$
ただし，S は標本空間である．

2. 2 次元空間の中にある事象とは (x, y) の集まった部分集合である．事象 A に対して
$$P\{(X, Y) \in A\} = \int \int_{(x, y) \in A} f(x, y)\, dx\, dy$$

$\boxed{\text{連続型確率変数の周辺分布}}$

X と Y の周辺確率分布を

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

で定義する．

$\boxed{X \text{ と } Y \text{ の共分散}}$

$$COV[X,\,Y] = E[(X-\mu_X)(Y-\mu_Y)], \qquad \text{ただし} \quad \mu_X = E[X],\,\mu_Y = E[Y]$$

で定義する．すると

$$V[X+Y] = V[X] + V[Y] + 2\,COV[X,\,Y]$$

$\boxed{X \text{ と } Y \text{ の相関係数}}$

$$\rho_{X,Y} = \frac{COV[X,\,Y]}{\sqrt{V[X]}\sqrt{V[Y]}}$$

$\boxed{\text{相関係数の性質}}$

1.
$$-1 \le \rho_{X,Y} \le 1$$
2. X と Y は無相関 $\Longleftrightarrow \rho_{X,Y} = 0$ （すなわち，$COV[X,\,Y] = 0$ ）

$\boxed{\text{独立な確率変数}}$

X と Y の同時確率分布において，あらゆる $x,\,y$ について

$$f(x,\,y) = g(x)\,\cdot\,h(y)$$

が成立するならば，X と Y は互いに独立であるという．ただし，$f(x,\,y)$ は X と Y の同時確率（密度）関数であり，$g(x)$ と $h(y)$ は X と Y の周辺確率（密度）関数である．

$\boxed{\text{性質}}$

1. $E(XY) = E(X)E(Y)$
2. 独立 $\Rightarrow$ 無相関．なぜならば，$COV(X,\,Y) = E(XY) - E(X)E(Y) = 0$
3. $V(X \pm Y) = V(X) + V(Y)$

$\boxed{n \text{ 個の確率変数の独立性}}$

n 個の確率変数 $X_1,\, X_2,\, \ldots,\, X_n$ が独立であるとは，あらゆる $x_1,\, x_2,\, \ldots,\, x_n$ に対して

$$f(x_1,\, x_2,\, \ldots,\, x_n) = f_1(x_1)f_2(x_2) \times \cdots \times f_n(x_n)$$

が成立することである．ただし，$f(x_1,\, x_2,\, \ldots,\, x_n)$ は $X_1,\, X_2,\, \ldots,\, X_n$ の同時確率（密度）関数であり，$f_i(x_i),\, i = 1,\, 2,\, \ldots,\, n,$ は X_i の周辺確率（密度）関数である．この場合，ふたつの確率変数 X_k と $X_l\,(1 \leq k \neq l \leq n)$ は独立である．

2004 年 6 月 29 日

総合科目　数理・自然 6 「統計学入門」講義レジュメ　(今野)

独立な確率変数の和

X と Y を独立な確率変数とする

1. $E[X+Y] = E[X] + E[Y]$
2. $V[X+Y] = V[X] + V[Y]$

$X_1, X_2, \ldots, X_n$ を互いに独立な確率変数

1. $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$
2. $V[X_1 + X_2 + \cdots + X_n] = V[X_1] + V[X_2] + \cdots + V[X_n]$

$X_1, X_2, \ldots, X_n$ を互いに独立に同一分布に従う場合

これらの平均値，分散を $E[X_1] = \mu, V[X_1] = \sigma^2$ とする．

1. $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \mathbb{E}[X_2] + \cdots + E[X_n] = n\mu$
2. $V[X_1 + X_2 + \cdots + X_n] = V[X_1] + V[X_2] + \cdots + V[X_n] = n\sigma^2$

正規分布の再帰性

$X_1, X_2, \ldots, X_n$ を独立同一に正規分布 $N(\mu, \sigma^2)$ に従うとする．

1. $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] = n\mu$
2. $V[X_1 + X_2 + \cdots + X_n] = V[X_1] + V[X_2] + \cdots + V[X_n] = n\sigma^2$
3. $X_1 + X_2 + \cdots + X_n$ は $N(n\mu, n\sigma^2)$ に従う．

大数の法則

$X_1, X_2, \ldots, X_n$ は互いに独立に同一分布に従い，その平均は $E[X_1] = \mu$ とする．n が大きいとき，

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

は μ に非常に近い．したがって，n が大きいときには $\bar{X}_n$ の値を μ とみなしてよい..

中心極限定理

$X_1, X_2, \ldots, X_n$ は互いに独立に同一分布に従い，その平均は
$E[X_1] = \mu, V[X_1] = \sigma^2$ とする．n が大きいとき，

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \quad \text{は } N(\mu, \tfrac{\sigma^2}{n}) \text{ で近似できる}$$

または

$$S_n = X_1 + X_2 + \cdots + X_n \qquad \text{は } N(n\mu,\, n\sigma^2) \text{ で近似できる}$$

$\boxed{\text{中心極限定理の応用}}$

4 万回コインを投げて 20300 回以上，または 19700 回以下表がでてることはどのくらいの確率であろうか?

$i\,(i = 1, 2, \ldots, 40000)$ 回目で表が出たとき，$X_i = 1$ とし，裏が出たとき，$X_i = 0$ とする．$40000(= n)$ 回中に表が出た総数 $S_{40000} = X_1 + X_2 + \cdots + X_{40000}$ の確率分布を求めればよい．このために，中心極限定理を利用する．$\mu = E[X_i] = 1/2$，$VAR[X_i] = 1/4$ なので，$n\mu = 40000 \times (1/2) = 20000$, $\sqrt{n}\sigma = \sqrt{40000} \times \sqrt{1/4} = 100$ となる．したがって

$$\begin{aligned}
&P(19700 \le X_1 + X_2 + \cdots + X_{40000} \le 20300) \\
&\quad = P\left(\frac{19700 - 20000}{100} \le \frac{S_{40000} - 20000}{100} \le \frac{20300 - 20000}{100}\right) \\
&\quad = P\left(-3 \le \frac{S_{40000} - 20000}{100} \le 3\right) \\
&\quad = \Phi(3) - \Phi(-3) = 2Q(3) = 2 \times 0.00135 = 0.0027
\end{aligned}$$

となる．ここで，$\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt$ と $Q(x) = 1 - \Phi(x)$, $(x \ge 0)$ である．