統計学 01

早稲田大学政治経済学部

第5回

西郷 浩

本日の目標

- 第7章 多次元の確率分布
  - 7.1 同時確率分布と周辺確率分布
  - 7.2 条件つき確率分布と独立な確率変数
  - 7.4 独立な確率変数の和
  - 説明は離散分布にしたがう確率変数にもとづいておこなう。

同時確率分布と周辺確率分布(1)

- 以下の試行を考える。
  - コインをひとつ投げる。
    - X: 表の枚数
  - つぎに $(X+1)$枚のコインを投げる。
    - Y: 表の枚数
- 2つの確率変数 (X, Y) の出方はどのようになるか。
  - $P(X=x, Y=y)$
    $=P(X=x)P(Y=y|X=x)$
    $=(1/2)\ {}_{x+1}C_y\ (1/2)^y (1/2)^{x+1-y}$

表1: (X, Y) の出現確率

y \ x	0	1	計
0	1/4	1/8	3/8
1	1/4	2/8	4/8
2	0	1/8	1/8
計	1/2	1/2	1

同時確率分布と周辺確率分布(2)

- 同時確率分布　$f(x,y)=P(X=x,Y=y)$
- 周辺確率分布　$g(x)=P(X=x)=\sum_{y} f(x,y)$

$$h(y)=P(Y=y)=\sum_{x} f(x,y)$$

- 同時確率分布における期待値

$$E\{\phi(X,Y)\}=\sum_{x}\sum_{y} f(x,y)\phi(x,y)$$

- 共分散

$$Cov(X,Y)=E\{(X-\mu_X)(Y-\mu_Y)\}=E(XY)-\mu_X\mu_Y$$

同時確率分布と周辺確率分布(3)

- 相関係数 $\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$
- 共分散の符号と (X, Y) の出方
  - Cov(X, Y) > 0 ⇒「X が大→ Y が大」となりやすい。
  - Cov(X, Y) < 0 ⇒「X が大→ Y が小」となりやすい。
    - 例：表1の確率分布 Cov(X,Y) = 1/8
- 相関係数の値： $-1\leq\rho_{XY}\leq 1$
  - ρ_{XY} が 1 に近い。⇒ $Y = a + bX$ (b>0) に近い。
  - ρ_{XY} が -1 に近い。⇒ $Y = a + bX$ (b<0) に近い。
  - ρ_{XY} が 0 に近い。⇒ 無相関
    - 例：表1の確率分布 $\rho_{XY}=1/\sqrt{7}=0.38$

同時確率分布と周辺確率分布(4)

- 確率変数の和･差の分散

$$V(X \pm Y) = V(X) \pm 2Cov(X, Y) + V(Y)$$

  - X と Y とに相関がない場合（ρ_{XY} =0⇔Cov(X,Y)=0）

$$V(X \pm Y) = V(X) + V(Y)$$

    - 2つの確率変数が独立（後述）であれば、
その2つの確率変数は無相関になる。
（逆は必ずしも成立しない。）
したがって、独立な2つの確率変数については、
「和の分散は、分散の和になる」が成立する。

条件つき確率分布

- 条件つき確率分布

$$g(x\mid y)=P(X=x\mid Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}=\frac{f(x,y)}{h(y)}$$

$$h(y\mid x)=P(Y=y\mid X=x)=\frac{P(X=x,Y=y)}{P(X=x)}=\frac{f(x,y)}{g(x)}$$

  - 例：表1の確率分布
    - $g(x|y)$を求める：
      - $g(0|0)=(1/4)/(3/8)=2/3$, $g(1|0)=(1/8)/(3/8)=1/3$; $g(0|0)+g(1|0)=1$
      - $g(0|1)=(1/4)/(4/8)=1/2$, $g(1|1)=(2/8)/(4/8)=1/2$; $g(0|1)+g(1|1)=1$
      - $g(0|2)=0/(1/8)=0$, $g(1|2)=(1/8)/(1/8)=1$; $g(0|2)+g(1|2)=1$

独立な確率変数(1)

- 2つの確率変数の独立性：任意の x, y について

$$f(x,y) = g(x)h(y) = P(X = x)P(Y = y)$$
$$\Leftrightarrow g(x \mid y) = g(x) \Leftrightarrow h(y \mid x) = h(y)$$

- 表1の場合、(X, Y) は独立ではない：
  - $f(1,2) = 1/8 \neq 1/2 \times 1/8 = g(1)\, h(2)$
    - $f(0,1) = 1/4 = 1/2 \times 4/8 = g(0) \times h(1)$ ではあるけれども、「どのような x, y についても」成り立たなければならない。
- 表7.5（教科書 p. 143）の場合、(X, Y) は独立である。
  - 2つのサイコロの出目 (X, Y)：
    - $f(x,y) = 1/36 = 1/6 \times 1/6 = g(x)\, h(y)$

独立な確率変数(2)

- 直観的には:
  - 「X の出方とY の出方が無関係である」
- 2つの確率変数が独立であれば、

$$E(XY) = E(X)E(Y)$$

  - このことから、

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0$$

つまり、「2つの確率変数が独立」
⇒「共分散が0」⇔「相関がない」
が成り立つ。

独立な確率変数の和(1)

- n 個の確率変数 X_i (i=1,2,…,n)の独立性

$$f(x_1, x_2, \ldots, x_n) = g_1(x_1)g_2(x_2)\cdots g_n(x_n)$$

- n 個の(独立とは限らない)確率変数の和の期待値

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

- n 個の独立な確率変数の和の分散

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

独立な確率変数の和(2)

- X_i が独立に同一の分布にしたがうなら、期待値・分散がすべての i について同じになるので、

$$E\left(\sum_{i=1}^{n} X_i\right) = n\mu \quad \text{ただし、} E(X_i) = \mu \ (i = 1, 2, \ldots, n)$$

$$V\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2 \quad \text{ただし、} V(X_i) = \sigma^2 \ (i = 1, 2, \ldots, n)$$

  - 例：ベルヌーイ確率変数の和

$$\text{ベルヌーイ確率変数} \quad X_i = \begin{cases} 1 & (\text{確率 } p) \\ 0 & (\text{確率 } 1 - p) \end{cases}$$

$$E(X_i) = 1 \times p + 0 \times (1 - p) = p$$

$$V(X_i) = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p)$$

独立な確率変数の和(3)

- X_i が相互に独立であれば、

$$X = \sum_{i=1}^{n} X_i \sim \mathrm{Bi}(n,\ p)$$

- したがって、

$$E(X) = np$$

$$V(X) = np(1-p)$$

- おまけ:二項分布の再生性

$$\mathrm{Bi}(n, p):\ X = \sum\nolimits_{i=1}^{n} X_i \quad \text{ただし} \quad X_i \sim \mathrm{Bernoulli}(p)$$

$$\mathrm{Bi}(m, p):\ Y = \sum\nolimits_{j=1}^{m} Y_j \quad \text{ただし} \quad Y_j \sim \mathrm{Bernoulli}(p)$$

$$Z = X + Y = \sum\nolimits_{i=1}^{n} X_i + \sum\nolimits_{j=1}^{m} Y_j \sim \mathrm{Bi}(n+m, p)$$

独立な確率変数の和(4)

- 応用:(標本)平均の期待値と分散

標本平均：$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ (ただし、X_i は相互に独立)

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)$$

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right)$$

とくに、X_i が相互に独立に同一の分布にしたがうなら

$E\left(\overline{X}\right) = \mu$　ただし　$E\left(X_i\right) = \mu$

$V\left(\overline{X}\right) = \sigma^2/n$　ただし　$V\left(X_i\right) = \sigma^2$

独立な確率変数の和(5)

- **正規確率変数の和**

$X_i \sim N\left(\mu_i, \sigma_i^2\right) \quad \left(i = 1, 2, \ldots, n\right)$

このとき、$W = \sum_{i=1}^{n} c_i X_i$ も正規分布にしたがう（証明略）。

$$E\left(W\right) = E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} E\left(c_i X_i\right) = \sum_{i=1}^{n} c_i E\left(X_i\right) = \sum_{i=1}^{n} c_i \mu_i$$

もし、X_i が相互に独立であれば、

$$V\left(W\right) = V\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} V\left(c_i X_i\right) = \sum_{i=1}^{n} c_i^2 V\left(X_i\right) = \sum_{i=1}^{n} c_i^2 \sigma_i^2$$

したがって、 $W \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$

独立な確率変数の和(6)

- とくに、相互に独立に同一の正規分布にしたがい、かつ、$c_i = 1/n$ のとき、W は標本平均となり、

$$X_i \sim N\left(\mu, \sigma^2\right)$$

$$W = \sum_{i=1}^{n} c_i X_i = \sum_{i=1}^{n} \left(\frac{1}{n}\right) X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

$$\bar{X} = W \sim N\left(\mu, \sigma^2/n\right)$$

独立な確率変数の和(7)

- 同一の分布から独立に発生した確率変数の平均値について、
  - その期待値は もとの確率変数の期待値に等しい、
  - その分散は もとの確率変数の 1/*n* 倍に等しい、

ということは、統計的推測に利用される重要な性質である。

- 「平均を取ることによってノイズを消す」ということは、統計の基本操作のひとつである。