多次元の確率分布と独立性

樋口さぶろお　https://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習　L07(2020-11-09 Mon)

最終更新: Time-stamp: "2020-11-06 Fri 16:32 JST hig"

今日の目標

- 同時分布から周辺分布, 母期待値, 母共分散, 母相関係数が計算できる 岩薩林 確率・統計 §3.3
- 確率変数の独立性を判定し利用できる 岩薩林 確率・統計 §3.3

L06-Q1

Quiz 解答:離散的な確率変数の母平均・母分散・母標準偏差

1. 母期待値 $\mathrm{E}[\mathrm{e}^X] = \frac{4}{12}\cdot\mathrm{e}^{-1} + \frac{5}{12}\cdot\mathrm{e}^{0} + \frac{3}{12}\cdot\mathrm{e}^{2}$.
2. 母平均値 $\mathrm{E}[X] = \frac{4}{12}\cdot(-1) + \frac{5}{12}\cdot 0 + \frac{3}{12}\cdot 2 = \frac{1}{6}(=\mu)$.
3. 母分散 $\mathrm{E}[X^2] = (-1)^2\cdot\frac{4}{12} + 0^2\cdot\frac{5}{12} + 2^2\frac{3}{12} = \frac{4}{3}$.
   $\mathrm{V}[X] \overset{\mathrm{V1}}{=} \mathrm{E}[X^2] - \mu^2 = \frac{47}{36}$.
   別解:
   $\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2] = \frac{4}{12}\cdot(-1-\frac{1}{6})^2 + \frac{5}{12}\cdot(0-\frac{1}{6})^2 + \frac{3}{12}(2-\frac{1}{6})^2 = \frac{47}{36}$.
4. 母標準偏差 $\sqrt{\mathrm{V}[X]} = \sqrt{\frac{47}{36}}$.
5. 確率 $\mathrm{E}[\mathrm{I}_{[X\le 1]}(X)] = \frac{4}{12}\cdot 1 + \frac{5}{12}\cdot 1 + \frac{3}{12}\cdot 0 = \frac{9}{12} = \frac{3}{4}$.

L06-Q2

Quiz 解答:確率変数の変換

$\mathrm{E}[X^2] \overset{\mathrm{V1}}{=} \mathrm{V}[X] + \mathrm{E}[X]^2 = 13$.

1. $\mathrm{E}[-X^2+2X-3] \overset{\text{E2}}{=} \mathrm{E}[-X^2]+\mathrm{E}[2X]+\mathrm{E}[-3] \overset{\text{E3}}{=}$
$-\mathrm{E}[X^2]+2\mathrm{E}[X]-3\mathrm{E}[1] \overset{\text{E1}}{=} -13+2\cdot 2-3\cdot 1=-12.$
2. $\mathrm{V}[-2X-3] \overset{\text{V2}}{=} (-2)^2\mathrm{V}[X]=36.$

ここまで来たよ

2 つの離散型確率変数の同時確率分布 高校 数学 B 岩薩林 確率・統計 p.56

例 6 枚のカードから無作為に 1 枚引く. ♡7 ♡8 ♡9 ◇8 ♠9 ♣9

2 つの離散型確率変数の同時分布

$X=$ 数, $Y=0$(赤札), 1(黒札) とすると (x,y) を得る確率は 2 変数の確率関数で書ける. 同時分布, 結合分布, joint distribution という.

$$p(x,y)=\begin{cases}\frac{1}{3} & ((x,y)=(8,0))\\ \frac{1}{6} & ((x,y)=(9,0))\\ \frac{1}{3} & ((x,y)=(9,1))\\ \frac{1}{6} & ((x,y)=(7,0))\\ 0 & (\text{他})\end{cases}$$

表で書いた方がまし. ここでは,「他」は省略.

岩薩林 確率・統計 では, p_{xy} とも.

$P(X=x,Y=y)$ とも. P(条件) 型の表記.

$y\backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

周辺分布

確率変数の周辺分布

同時分布 $p(x,y)$ に対して,
X の周辺分布 $p_X(x)$, Y の周辺分布 $p_Y(y)$ は,

$$p_X(x) = p(x,\bullet) = \sum_y p(x,y), \quad p_Y(y) = p(\bullet,y) = \sum_x p(x,y)$$

要するに 一方を無視した分布. 小計.

岩薩林 確率・統計 例題 3.4(p.57)

$y \backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

同時分布が与えられたときの母期待値は，周辺分布から楽に計算

同時分布が与えられたときの母期待値

岩薩林 確率・統計 (3.16)p.60

$$\mathrm{E}[g(X,Y)] \overset{\text{定義}}{=} \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} g(x,y) \cdot p(x,y)$$

同時分布が与えられたときの確率

条件 $a(X,Y)$ で定まる事象 $\{(x,y)|a(x,y)\} \subset \Omega$ に対して，

$$P(a(X,Y)) = \mathrm{E}[\mathrm{I}_{[a(X,Y)]}(X,Y)] = \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} \mathrm{I}_{[a(X,Y)]}(x,y) \cdot p(x,y)$$

(P2)

L07-Q1

Quiz(多変数の確率変数の期待値)

2 変数 X, Y の離散型確率分布を考える．同時分布 $p(x, y)$ が下の表で与えられる．

$y \backslash x$	1	2	3
0	0	$\frac{2}{12}$	$\frac{1}{12}$
2	$\frac{4}{12}$	0	$\frac{5}{12}$

1. 母期待値 $\mathrm{E}[X^2 + e^Y]$ を求めよう．
2. 母期待値 $\mathrm{E}[\mathrm{I}_{[XY \geq 2]}(X, Y)]$ を求めよう．
3. 周辺分布 $p_{\mathrm{X}}(x)$, $p_{\mathrm{Y}}(y)$ を求めよう．

復習:先週知った, 確率変数 X の母期待値, 母分散についての公式

岩薩林 確率・統計 重要事項のまとめ p. xii 参照 (En, Vn の番号は載ってないけど)

2 次元の確率分布の母期待値の性質 高校 数学 B 岩薩林 確率・統計 (3.17)p.61

$$\mathrm{E}[g_1(X,Y)+g_2(X,Y)] = \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)] \tag{E5}$$

$$\text{特に } \mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \tag{E6,(3.17)}$$

証明

$$\begin{aligned}\mathrm{E}[g_1(X,Y)+g_2(X,Y)] &= \sum_x \sum_y (g_1(x,y)+g_2(x,y))p(x,y)\\ &= \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)].\end{aligned}$$

X だけ, Y だけの関数の母期待値

x だけ, y だけの関数の母期待値は,
下の左辺= **同時分布** で計算しても
下の右辺= **周辺分布** で計算しても
同じ結果.

$$\mathrm{E}[g(X)] = \sum_x \sum_y g(x) \cdot p(x,y) = \sum_x g(x) \sum_y p(x,y) = \sum_x g(x) \cdot p_{\mathrm{X}}(x)$$

$$\mathrm{E}[g(Y)] = \sum_y \sum_x g(y) \cdot p(x,y) = \sum_y g(y) \sum_x p(x,y) = \sum_y g(y) \cdot p_{\mathrm{Y}}(y)$$

Example (2 次元の確率分布の母期待値のよくある間違い)

$$\mathrm{E}[g(X,Y)] \neq g(\mathrm{E}[X],\mathrm{E}[Y])$$

だって, $\sin(\log_x y) \neq \log_{\sin x}(\sin y)$ じゃん.
特に $\mathrm{E}[XY] \neq \mathrm{E}[X]\mathrm{E}[Y]$.
$\times$ は, $+$ とは別. (独立でないときは)

Example (2 次元の確率分布の母分散のよくある間違い)

$$\mathrm{V}[X+Y] \neq \mathrm{V}[X]+\mathrm{V}[Y]$$

母分散は, 母期待値とは別. (独立でないときは)

ここまで来たよ

母共分散高校 数学 B

母共分散 covariance 岩薩林 確率・統計 (3.19)p.61,(3.20)p.62

X, Y が確率変数で, $\mu_{\mathrm{X}} = \mathrm{E}[X], \mu_{\mathrm{Y}} = \mathrm{E}[Y]$ とおいたとき,

$$\text{母共分散}\ \mathrm{Cov}[X,Y] \overset{\text{定義}}{=} \mathrm{E}[(X-\mu_{\mathrm{X}})(Y-\mu_{\mathrm{Y}})] \tag{3.19}$$

$$= \boxed{\text{岩薩林 確率・統計 (3.20)}} \cdots = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y]. \tag{C1,(3.20)}$$

母相関係数 correlation 岩薩林 確率・統計 (3.22)p.64

$$\text{母相関係数}\rho[X,Y] \overset{\text{定義}}{=} \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{V}[X]}\sqrt{\mathrm{V}[Y]}} \tag{3.22}$$

母共分散, 母相関係数のの性質

$$-1 \le \rho[X,Y] \le 1 \tag{3.23}$$

$$\mathrm{Cov}[aX+b, cY+d] = ac \cdot \mathrm{Cov}[X,Y]. \tag{C2}$$

$$\rho[aX+b, cY+d] = \frac{ac}{|a||c|}\rho[X,Y] \tag{C3}$$

L07-Q2

Quiz(独立と限らない確率変数の母期待値)

確率変数 X, Y を考える.
$\mathrm{E}[X] = 2, \mathrm{E}[Y] = 3, \mathrm{V}[X] = 5, \mathrm{V}[Y] = 11, \mathrm{Cov}[X, Y] = 7$ である.

1. $\mathrm{E}[-2X + 3Y]$ を求めよう.
2. $\mathrm{V}[-2X + 3Y]$ を求めよう.

ここまで来たよ

6 離散型確率変数

7 多次元の確率分布と独立性
- 2 次元の確率分布
- 母共分散
- 確率変数の独立性

確率変数の独立性高校 数学 B 岩薩林 確率・統計 (3.14)p.58

独立性

確率変数 X, Y が同時分布 $p(x, y)$ を持つとき, X, Y が独立 (independent) とは, 次の関係が成立することをいう (世の中には, 同値な定義が多数).

$$p(x, y) = p_{\mathrm{X}}(x) \times p_{\mathrm{Y}}(y)$$

X, Y が独立とは,

X, Y が互いに「無関係」であること

同時分布が, 周辺分布だけから (積で) 決まっちゃうこと

事象 A, B が独立 $\Leftrightarrow P(A \text{ かつ } B) = P(A) \times P(B)$ 岩薩林 確率・統計 p.40, を確率変数に対して書いたもの.

L07-Q3

Quiz(2 つの離散型確率変数の母期待値・母平均値・母共分散・確率・独立性)

離散型確率変数 X, Y の同時分布は次の表で与えられる.

$y \backslash x$	1	3
2	$\frac{1}{7}$	$\frac{2}{7}$
4	0	$\frac{4}{7}$

で与えられる.

1. X, Y が独立かどうか判定しよう.
2. 母平均値 $\mathrm{E}[X], \mathrm{E}[Y]$ を求めよう.
3. 母期待値 $\mathrm{E}[XY], \mathrm{Cov}[X, Y]$ を求めよう.

L07-Q4

Quiz(2 つの独立な離散型確率変数の母期待値・母平均値・母共分散・確率)

離散型確率変数 X, Y の同時分布は次の表で与えられる.

$y\backslash x$	1	3
2	1/12	3/12
4	2/12	6/12

で与えられる.

1. X, Y が独立かどうか判定しよう.
2. 母平均値 $\mathrm{E}[X], \mathrm{E}[Y]$ を求めよう.
3. 母期待値 $\mathrm{E}[XY], \mathrm{Cov}[X, Y]$ を求めよう.
4. 母期待値 $\mathrm{E}[X^2], \mathrm{E}[X^2Y]$ を求めよう.

岩薩林 確率・統計 例題 3.4(p.57)

X, Y が独立，はラッキー

$x = 1, 2, \cdots, 100, y = 1, 2, \cdots, 200$ の確率分布を暗記しろって言われたときに，

- 独立じゃなかったら，100×200 個の数をおぼえなきゃいけない
- 独立なら $100 + 200$ 個の数だけおぼえればいい．

X, Y が独立なときに成立するとてもいい性質 岩薩林 確率・統計 (3.18)p.61

$$\mathrm{E}[g_1(X) \times g_2(Y)] = \mathrm{E}[g_1(X)] \times \mathrm{E}[g_2(Y)] \quad \text{(IE1)}$$

$$\text{特に } \mathrm{E}[XY] = \mathrm{E}[X] \times \mathrm{E}[Y] \quad \text{(IE2,(3.18))}$$

$$\text{特に } \mathrm{Cov}[X, Y] = 0 \quad \text{(IC1,(3.21))}$$

岩薩林 確率・統計 (3.21)p.63

$$\mathrm{V}[g_1(X) + g_2(Y)] = \mathrm{V}[g_1(X)] + \mathrm{V}[g_2(Y)] \quad \text{(IC2)}$$

岩薩林 確率・統計 (3.21)p.63

$$\mathrm{V}[X + Y] = \mathrm{V}[X] + \mathrm{V}[Y] \quad \text{(IC3,(3.21))}$$

岩薩林 確率・統計 (3.21)p.63

$$\mathrm{V}[aX + bY + c] = a^2\mathrm{V}[X] + b^2\mathrm{V}[Y] \quad \text{(IC4)}$$

IC3,IC4 は, 独立でなくても $\mathrm{Cov}[X, Y] = 0$ だけで成り立つ.

$\mathrm{Cov}[X, Y] = 0$ は独立の 必要条件 (だが 十分条件 ではない)

一部の証明

$$\begin{aligned}
\mathrm{E}[XY] &= \sum_x \sum_y xy \cdot p(x,y) \\
&\overset{\text{独立}}{=} \sum_x \sum_y xy \cdot p_\mathrm{X}(x) \times p_\mathrm{Y}(y) \\
&= \sum_x x \cdot p_\mathrm{X}(x) \times \sum_y y \cdot p_\mathrm{Y}(y) = \mathrm{E}[X] \times \mathrm{E}[Y]
\end{aligned}$$

$$\begin{aligned}
\mathrm{V}[X+Y] &\overset{\text{V1}}{=} \mathrm{E}[(X+Y)^2] - \mathrm{E}[X+Y]^2 \\
&\overset{\text{E6}}{=} \mathrm{E}[X^2] + 2\mathrm{E}[XY] + \mathrm{E}[Y^2] - (\mathrm{E}[X]^2 + 2\mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}[Y]^2) \\
&\overset{\text{V1,C1}}{=} \mathrm{V}[X] + 2\mathrm{Cov}[X,Y] + \mathrm{V}[Y] \\
&\overset{\text{IC1}}{=} \mathrm{V}[X] + 0 + \mathrm{V}[Y]
\end{aligned}$$

L07-Q5

Quiz(独立な確率変数の母期待値)

独立な確率変数 X, Y を考える．
$\mathrm{E}[X] = 2, \mathrm{E}[Y] = 3, \mathrm{V}[X] = 5, \mathrm{V}[Y] = 11$ である．

1. $\mathrm{E}[(-2X + 3Y)(X + 5Y)]$ を求めよう．
2. $\mathrm{V}[-2X + 3Y]$ を求めよう．

岩薩林 確率・統計 問題 6(p.64)

岩薩林 確率・統計 第 3 章練習問題 2(p.73)

岩薩林 確率・統計 第 3 章練習問題 5(p.73)