多次元の確率変数

樋口さぶろお　https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

確率統計 I　L04(2021-04-28 Wed)

最終更新: Time-stamp: "2021-04-27 Tue 09:55 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

L03-Q1
Quiz 解答:連続型一様分布

1. $\mathrm{E}[X^k] = \frac{1}{d-c}\int_c^d x^k \, \mathrm{d}x = \frac{1}{k+1}\frac{d^{k+1}-c^{k+1}}{d-c} = \frac{1}{k+1}(d^k + d^{k-1}c + \cdots + c^k)$.
2. $\mathrm{E}[X^1] = \frac{c+d}{2}$.
3. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{(d-c)^2}{12}$. $\sqrt{\mathrm{V}[X]} = \frac{d-c}{\sqrt{12}} \simeq \frac{d-c}{3.5}$.

L03-Q2
Quiz 解答:連続型一様分布の応用

1. $X \sim \mathrm{U}(18, 20)$.
2. $\mathrm{E}[X^1] = 19$ [mm].
3. $\mathrm{E}[6X^2] = 2168 \neq 6 \cdot 19^2 = 2166$ [mm^2].
4. $P(6X^2 \leq 2000) = P(X \leq (\frac{2000}{6})^{1/2}) = \int_1 9^{(2000/6)^{1/2}} \frac{1}{20-18} \, \mathrm{d}x = 5\sqrt{10/3} - 9$.
5. $\mathrm{E}[X^3] = 6878 \neq 19^3 = 6859$ [mm^3].

L03-Q3

Quiz 解答:確率変数の変換

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

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. $\mathrm{V}[-2X - 3] \overset{\text{V2}}{=} (-2)^2 \mathrm{V}[X] = 36.$

L03-Q4

Quiz 解答:連続型一様分布の母期待値

$\mathrm{E}[X] = \frac{-1+1}{2} = 0, \mathrm{V}[X] = \frac{(1-(-1))^2}{12} = \frac{1}{3}$

1. $\mathrm{E}[2X + 4] = 2 \cdot 0 + 4.$
2. $\mathrm{V}[2X + 4] = 2^2 \times \frac{1}{3}.$
3. $2X + 4 > 5 \Leftrightarrow X > \frac{1}{2}.\ P(X > \frac{1}{2}) = \int_{\frac{1}{2}}^{1} \frac{1}{2}\,\mathrm{d}x = \frac{1-\frac{1}{2}}{2} = \frac{1}{4}.$

別解 $Y = 2X + 4 \sim \mathrm{U}(2,6)$ なので,

1. $\mathrm{E}[Y] = \frac{2+6}{2} = 4$.
2. $\mathrm{V}[Y] = \frac{(6-2)^2}{12} = \frac{1}{3}$.
3. 面積の比を考えて, $P(Y > 5) = \frac{6-5}{6-2} = \frac{1}{4}$.

L03-Q5

Quiz 解答:連続型一様分布

1.
$$f_{\mathrm{Z}}(z) = \frac{1}{2\sqrt{3}} \begin{cases} 1 & (-\sqrt{3} \le y < \sqrt{3}) \\ 0 & (他) \end{cases}$$

   より,
$$f_{\mathrm{X}}(x) = \frac{1}{2\sqrt{3}} \begin{cases} 1 & (1 - 2\sqrt{3} \le y < 1 + 2\sqrt{3}) \\ 0 & (他) \end{cases}.$$

すなわち, $X \sim \mathrm{U}(1-2\sqrt{3}, 1+2\sqrt{3})$.

2. $\mathrm{E}[X] = \frac{(1-2\sqrt{3})+(1+2\sqrt{3})}{2}$, または, $\mathrm{E}[X] = 2\mathrm{E}[Z] + 1 = 0 + 1$.
3. $\mathrm{V}[X] = \frac{((1+2\sqrt{3})-(1-2\sqrt{3}))^2}{12}$, または, $\mathrm{V}[X] = 2^2\mathrm{V}[Z] = 2^2 \cdot 1$.

ここまで来たよ

3 連続型一様分布, 変数変換, 確率分布の標準化

3 多次元の確率変数
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

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 $p_\mathrm{X}(x)$, Y の周辺分布 $p_\mathrm{Y}(y)$ は,

$$p_\mathrm{X}(x) = p(x,\bullet) = \sum_y p(x,y), \quad p_\mathrm{Y}(y) = p(\bullet,y) = \sum_x p(x,y)$$

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，

$$P(a(X,Y)) = \mathrm{E}[\mathrm{I}_{[a(X,Y)]}(X,Y)] = \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} \mathrm{I}_{[a(X,Y)]}(x,y) \cdot p(x,y) \tag{P2}$$

L04-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

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

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

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

3 連続型一様分布, 変数変換, 確率分布の標準化

3 多次元の確率変数
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

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

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

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

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

L04-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.

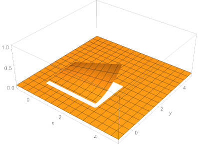
ここまで来たよ

3. 連続型一様分布, 変数変換, 確率分布の標準化

3. 多次元の確率変数
   - 2 次元の確率分布
   - 母共分散
   - 多次元の連続型確率分布

多次元の連続型確率分布 岩薩林 確率・統計 §4.6

2 次元の同時確率密度関数 $f(x, y)$.
$f_{\mathrm{XY}}(x, y) \geq 0$, 大きいほど, その (x, y) が「出やすい」



確率変数の母期待値の定義 岩薩林 確率・統計 SS3.3,4.6

$$\text{離散型} \quad \mathrm{E}[g(X, Y)] = \sum_{x} \sum_{y} g(x, y) \cdot p(x, y) \tag{3.16}$$

$$\text{連続型} \quad \mathrm{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) \cdot f(x, y)\, \mathrm{d}x\mathrm{d}y \tag{4.32}$$

確率は $\mathrm{E}[\mathrm{I}_{[X,Y\text{ の条件}]}(x, y)] =$ その部分の体積.

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 $p_{\mathrm{X}}(x)$, Y の周辺分布 $p_{\mathrm{Y}}(y)$ は，

$$\text{離散型}\ p_{\mathrm{X}}(x) = p(x,\bullet) = \sum_y p(x,y), \quad p_{\mathrm{Y}}(y) = p(\bullet,y) = \sum_x p(x,y), \tag{3.11,3.12}$$

$$\text{連続型}\ f_{\mathrm{X}}(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y, \quad f_{\mathrm{Y}}(y) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x \tag{4.28,4.29}$$

要するに 一方を無視した分布．小計．

L04-Q3

Quiz(独立と限らない多次元の連続型確率変数)

連続型確率変数 X, Y の同時確率密度関数は

$$f(x,y) = \begin{cases} \frac{1}{24}(2+xy^2) & (0 \leq x \leq 3, 0 \leq y \leq 2) \\ 0 & (\text{他}) \end{cases}$$

で与えられる.

1. 母期待値 $\mathrm{E}[X^3Y]$ を求めよう.
2. 確率 $P(X \leq 2 \ \text{かつ} \ Y \leq 1)$ を求めよう.