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　L06(2022-05-16 Mon)

最終更新: Time-stamp: "2022-05-16 Mon 06:49 JST hig"

今日の目標

- 同時分布から周辺分布, 条件付き確率, 母期待値, 母共分散が計算できる 岩薩林 確率・統計 §3.3

L05-Q1

Quiz 解答: 指数分布の確率

確率密度関数は $f(x) = \begin{cases} 2\mathrm{e}^{-2x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$.

1. 確率 $P(X \leq 3) = \int_{-\infty}^{3} f(x)\mathrm{d}x = F(3) = 1 - \mathrm{e}^{-6}$.
2. $F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x = \begin{cases} 0 & (x \leq 0) \\ 1 - \mathrm{e}^{-2x} & (x > 0) \end{cases}$.
3. 確率 $P(3 < X \leq 4) = \int_{3}^{4} f(x)\mathrm{d}x = F(4) - F(3) = \mathrm{e}^{-6} - \mathrm{e}^{-8}$.
4. 確率 $P(X \geq 4) = \int_{4}^{+\infty} f(x)\mathrm{d}x = F(\infty) - F(4) = \mathrm{e}^{-8}$.
5. $F(a) = 3/4$ より，$F^{-1}(3/4)$. $2\mathrm{e}^{-2a} = 3/4$ より，$a = -\frac{1}{2}\log\frac{3}{8}$.

L05-Q3

Quiz 解答: 正規分布の確率

1. $\mathrm{E}[X] = 4$.

2. $\mathrm{V}[X] = 3^2$.
3. $x = 4$ を真ん中に幅 3 くらいの正規分布の確率密度関数のグラフ．標準正規分布を，横に 3 倍，縦に 1/3 倍，右に 4 平行移動.

L05-Q4

Quiz 解答: 正規分布の確率

$$f(x) = C' \mathrm{e}^{-\frac{(x-3)^2}{2(1/4)^2}}$$

とかけるので，$X \sim \mathrm{N}(3, (\frac{1}{4})^2)$ であり，

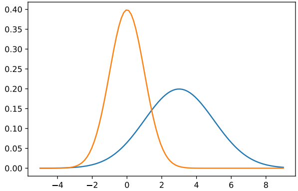
1. $\mathrm{E}[X] = 3$.
2. $\mathrm{V}[X] = (\frac{1}{4})^2$.
3. $x = 3$ を真ん中に幅 $\frac{1}{4}$ くらいの正規分布の確率密度関数のグラフ．標準正規分布を，横に 1/4 倍，縦に 4 倍，右に 3 平行移動.

L05-Q5

Quiz 解答: 正規分布の確率

X を標準化すると, $Z = \frac{X-3}{2}$.

1. $f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi\cdot 2^2}}\mathrm{e}^{-\frac{(x-3)^2}{2\cdot 2^2}}$, $f_{\mathrm{Z}}(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}$. X, Z の累積分布関数をそれぞれ $F_{\mathrm{X}}(x)$ =scipy.stats.norm(loc=2,scale=3).cdf(x), $F_{\mathrm{Z}}(z)$ =scipy.stats.norm(loc=0,scale=1).cdf(z) とかく.



2. $\mathrm{E}[X^2] = \mathrm{V}[X] + \mathrm{E}[X]^2 = 2^2 + 3^2$.

3. $P(X \geq 5) = P(5 \leq X < +\infty) = P(\frac{5-3}{2} \leq \frac{X-3}{2} < +\infty) = P(\frac{5-3}{2} \leq Z < +\infty) = 1 - F_{\mathrm{Z}}(1)$ =1-scipy.stats.norm(loc=0,scale=1).cdf(1) $(= 1 - F_{\mathrm{X}}(5)) = 0.15877.$
4. $P(1 \leq X \leq 7) = P(\frac{1-3}{2} \leq Z \leq \frac{7-3}{2}) = F_{\mathrm{Z}}(2) - F_{\mathrm{Z}}(-1)(= F_{\mathrm{X}}(7) - F_{\mathrm{X}}(1)) = 0.81859.$
5. $P(3 \leq X \leq 7) = P(\frac{3-3}{2} \leq Z \leq \frac{7-3}{2}) = F_{\mathrm{Z}}(z) - \frac{1}{2}(= F_{\mathrm{X}}(7) - F_{\mathrm{X}}(3)) = 0.47725.$
6. $a = F_{\mathrm{X}}^{-1}(3/4) =$ scipy.stats.norm(loc=2,scale=3).ppf(0.75)$= 4.0234.$

ここまで来たよ

6 指数分布・一般の正規分布

6 多次元の確率変数
- 2 次元の確率分布
- 母共分散
- 条件付き確率

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				

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,

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

L06-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. 確率 $P(XY \geq 2)$ を求めよう．
3. 周辺分布 $p_{\mathrm{X}}(x)$, $p_{\mathrm{Y}}(y)$ を求めよう．

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

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

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				

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.

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

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

L06-Q2

Quiz

さっきの問で

1. 周辺分布を求めよう.
2. 周辺分布から $\mathrm{E}[2X^2 + \mathrm{e}^Y]$ を求めよう.
3. 母分散 $\mathrm{V}[X]$ を求めよう.

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, $\mu_X = \mathrm{E}[X], \mu_Y = \mathrm{E}[Y]$ とおいたとき,

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

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

L06-Q3

母共分散

さっきの問で

1. 母期待値 $\mathrm{E}[XY]$ を求めよう.
2. 母共分散 $\mathrm{Cov}[X, Y]$ を求めよう.

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 $\mathrm{E}[X \times Y] \neq \mathrm{E}[X] \times \mathrm{E}[Y]$. $\times$ を, $+$ と区別しよう.

2 次元の確率分布の母分散のよくある間違い

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

母分散を, 母期待値と区別しよう.

L06-Q4

母分散

さっきの問で母分散 $\mathrm{V}[X+Y]$ を求めよう.

ここまで来たよ

条件付き確率 岩薩林 確率・統計 §2.3

条件付き確率の定義 岩薩林 確率・統計 (2.12)

条件 (事象) A のもとでの, 事象 B の条件付き確率

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

縦棒「|」の前後は不平等: P(確率を考える事象|条件の事象)
$P(B|A) \neq P(A|B), \neq P(A \cup B)$.
意味

A が起きる場合に限定した, B が起きる確率

縦棒の前だけ見る (縦棒の後ろを固定する) と, ただの確率

離散型確率変数に対する条件付き確率

事象 $B \leftrightarrow$ 事象 $Y = 50$
同時確率分布

$p(x, y) = P(X = x かつ Y = y) =$

$y \backslash x$	158	160	165
45	3/8	0	1/12
50	1/8	1/3	1/12

離散型確率変数に対する条件付き確率 岩薩林 確率・統計 (3.15)

条件 (事象)$Y = y$ のもとでの，事象 $X = x$ の条件付き確率の表し方
$P(X = x | Y = y) = p_{\mathrm{X|Y}}(x \text{ の値} | y \text{ の値})$

例 $p_{\mathrm{X|Y}}(158|45), p_{\mathrm{Y|X}}(45|158)$ 意味

自分の言葉でどうぞ

同時確率分布と周辺分布で書くと，

$$P(X = x | Y = y) = p_{\mathrm{X|Y}}(x|y) = \frac{p(x, y)}{p_{\mathrm{Y}}(y)}$$

$$P(Y = y | X = x) = p_{\mathrm{Y|X}}(y|x) = \frac{p(x, y)}{p_{\mathrm{X}}(x)}$$

縦棒の前だけ見る (縦棒の後ろを固定する) と，前の変数についての，ただの確率分布

L06-Q5

Quiz(条件付き分布)

次の 6 枚のカードから無作為に 1 枚のカードを引く.

♡7 ♡8 ♡9 ◇8 ♠9 ♣9

$X=$ 数, $Y=0$(赤札), 1(黒札) とすると同時分布は次のようになる.

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

1. 9 の札が出る条件のもとで赤札が出る, 条件付き確率を求めよう.
2. 赤札が出る条件のもとで 9 の札が出る, 条件付き確率を求めよう.

条件付き確率と周辺分布から，同時確率分布，周辺分布

同時確率分布，周辺分布，条件付き確率の関係

$$\begin{aligned} p_{\mathrm{XY}}(x,y) =& p_{\mathrm{X|Y}}(x|y) \times p_{\mathrm{Y}}(y) \\ =& p_{\mathrm{Y|X}}(y|x) \times p_{\mathrm{X}}(x). \end{aligned}$$

さらに周辺分布との関係 岩薩林 確率・統計 全確率の法則 (2.13)

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

$$p_{\mathrm{Y}}(y) = \sum_x p_{\mathrm{Y|X}}(y|x) \times p_{\mathrm{X}}(x)$$

(1) 条件付き確率の定義の分母をはらっただけ.
(2) いったん同時確率が求まれば，周辺分布など，ぜんぶ求まる.

L06-Q6

Quiz(同時分布・条件付き分布・周辺分布)

離散型確率変数 X は値 $2, 3$, Y は値 $20, 30$ をとる.
次がわかっている.

$$p_{\mathrm{Y}}(y) = \begin{cases} \frac{1}{3} & (y = 20) \\ \frac{2}{3} & (y = 30) \\ 0 & (\text{他}) \end{cases}$$

$$p_{\mathrm{X|Y}}(x|20) = \begin{cases} \frac{3}{4} & (x = 2) \\ \frac{1}{4} & (x = 3) \end{cases}, \quad p_{\mathrm{X|Y}}(x|30) = \begin{cases} \frac{1}{2} & (x = 2) \\ \frac{1}{2} & (x = 3) \end{cases}.$$

1. 同時分布を求めよう
2. X の周辺分布を求めよう.

岩薩林 確率・統計 例題 3.3,3.5,3 章練習問題 2