

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(2024-05-27 Mon)

最終更新: Time-stamp: "2024-05-27 Mon 07:03 JST hig"

今日の目標

- 多次元確率変数の同時分布から周辺分布, 母期待値, 母共分散を求められる。 [岩薩林 確率・統計 §3.3](#)
- 次の計算ができる: 同時確率分布, 周辺分布, 条件付き確率 ([岩薩林 確率・統計 §2.3, p.59](#), ベイズの定理 [岩薩林 確率・統計 p.42](#))



L05-Q1

Quiz 解答: 標準正規分布の確率

累積分布関数を $\Phi(z)$ とする. `rvz=stats.norm(loc=0,scale=1)` とする.

- ① $P(Z \leq 1.23) = \int_{-\infty}^{1.23} \phi(z) dz = \Phi(1.23) - \Phi(-\infty) = \Phi(1.23) - 0 = \text{rvz.cdf}(1.23) = 0.5708.$
- ② $P(-0.56 < Z \leq +1.23) = \int_{-0.56}^{1.23} \phi(z) dz = \Phi(1.23) - \Phi(-0.56) = \text{rvz.cdf}(1.23) - \text{rvz.cdf}(-0.56) = 0.6030.$
- ③ $0.025 = P(Z > d) = P(d < Z < +\infty) = \Phi(+\infty) - \Phi(d) = 1 - \Phi(d).$
よって, $\Phi(d) = 1 - 0.025.$
 $d = F^{-1}(1 - 0.025) = \text{rvz.ppf}(0.975) = 1.960.$

L05-Q2

Quiz 解答: 正規分布の確率

- ① $E[X] = 4.$

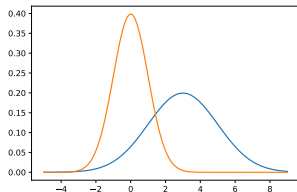
- ② $V[X] = 3^2$.
- ③ $x = 4$ を真ん中に幅 3 くらいの正規分布の確率密度関数のグラフ. 標準正規分布を, 横に 3 倍, 縦に $1/3$ 倍, 右に 4 平行移動.

L05-Q3

Quiz 解答: 正規分布の確率

X を標準化すると, $Z = \frac{X-3}{2}$.

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



- ② $E[X^2] = V[X] + E[X]^2 = 2^2 + 3^2$.
- ③ $P(X \geq 5) = P(5 \leq X < +\infty) = P(\frac{5-3}{2} \leq \frac{X-3}{2} < +\infty) = P(\frac{5-3}{2} \leq Z < +\infty) = 1 - \Phi(1) = 1 - \text{scipy.stats.norm(loc=0, scale=1).cdf(1)}$
 $(= 1 - F_X(5)) = 0.15877$.
- ④ $P(1 \leq X \leq 7) = P(\frac{1-3}{2} \leq Z \leq \frac{7-3}{2}) = \Phi(2) - \Phi(-1) = 0.81859$.
- ⑤ $P(3 \leq X \leq 9) = P(\frac{3-3}{2} \leq Z \leq \frac{9-3}{2})$.
- ⑥ $\frac{3}{4} = P(-\infty < X \leq d) = P(-\infty < Z \leq \frac{d-3}{2})$ より $d' = (d-3)/2$.
 $d' = \Phi^{-1}(\frac{3}{4}) = \text{scipy.stats.norm(loc=0, scale=1).ppf(0.75)}$.

ここまで来たよ

5 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.

コンマ前が X , 後が Y の値. 明示のため $p_{XY}(x, y)$ と書くことも.

岩薩林 確率・統計 では, p_{xy} とも. $P(X = x, Y = y)$ とも. $P(\text{条件})$ 型の表記.

表で書いた方がまし. ここでは, 「他」は省略.

表の縦軸横軸に意味なし. ラベル X, Y をよく見て.

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

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

今まで時々見た、「 $X = 2Y + 7$ 」のような式は、同時分布が次の形であることを意味する。

$y \backslash x$	7	8	9
0	p	0	0
1	0	0	$1 - p$

定義 (同時分布の母期待値 岩薩林 確率・統計 (3.16)p.60)

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

定義 (同時分布の確率)

条件 $a < X \leq b$ かつ $c < Y \leq d$ で定まる確率は、

$$P(a < X \leq b \text{ かつ } c < Y \leq d) = \sum_{a < x \leq b} \sum_{c < y \leq d} p(x, y) \quad (\text{P2})$$

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 $E[2X^2 + e^Y]$ を求めよう.
- ② 確率 $P(XY \geq 2)$ を求めよう.
- ③ 周辺分布 $p_X(x), p_Y(y)$ を求めよう.

命題 (多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)

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

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

証明

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

周辺分布 岩薩林 確率・統計 §3.3

定義 (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				

命題 (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.

$$E[g(X)] = \sum_x \sum_y g(x) \cdot p(x, y) = \sum_x g(x) \sum_y p(x, y) = \sum_x g(x) \cdot p_X(x)$$

$$E[g(Y)] = \sum_y \sum_x g(y) \cdot p(x, y) = \sum_y g(y) \sum_x p(x, y) = \sum_y g(y) \cdot p_Y(y)$$

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 $E[2X^2 + e^Y]$ を求めよう.
- ③ 母分散 $V[X]$ を求めよう.

ここまで来たよ

5 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

母共分散 高校 数学 B

定義 (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, $\mu_X = E[X], \mu_Y = E[Y]$ とおいたとき,

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

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

L06-Q3

母共分散

さっきの問で

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

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

だって, $\sin(\log_x y) \neq \log_{\sin x}(\sin y)$ じゃん.

特に $E[X \times Y] \neq E[X] \times E[Y]$. \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

$$V[X + Y] \neq V[X] + V[Y]$$

母分散を, 母期待値と区別しよう.

ここまで来たよ

5 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

条件付き確率 岩薩林 確率・統計 §2.3

定義 (条件付き確率 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「|」=given」の前後は不平等: $P(\text{確率を考える事象}|\text{条件の事象})$
 $P(B|A) \neq P(A|B), \neq P(A \cup B)$.

意味 A が起きる場合に限定した, B が起きる確率

$P(B|A)$ を $P(\text{of})B \text{ given } A$ と発音する人がいる. given 条件=条件が与えられたとき, という言い方から.

縦棒の前だけ見る (縦棒の後を固定する) と, ただの確率. 全事象は 1.

離散型確率変数に対する条件付き確率

同時確率分布

$$p(x, y) = P(X = x \text{かつ} Y = y) =$$

$y \backslash x$	150	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 $X = x$ の条件付き確率の記号

$$P(X = x | Y = y) = p_{X|Y}(x \text{の値} | y \text{の値})$$

例 $p_{X|Y}(150|45), p_{Y|X}(45|150)$. 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. 意味 自分の言葉でどうぞ)

命題 (条件付き確率を同時確率分布と周辺分布で表す式 岩薩林 確率・統計 (2.12))

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

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

L06-Q4

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				

- 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.
- 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

命題 (同時確率分布, 周辺分布, 条件付き確率の関係)

乗法公式 (同時確率分布との関係) 岩薩林 確率・統計 (2.13)

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

全確率の法則 (周辺分布との関係) 岩薩林 確率・統計 (2.17),(2.20)

$$\begin{aligned} p_X(x) &= \sum_y p_{X|Y}(x|y) \times p_Y(y), \\ p_Y(y) &= \sum_x p_{Y|X}(y|x) \times p_X(x) \end{aligned}$$

ベイズの定理 I

定理 (ベイズの定理 岩薩林 確率・統計 (2.18),(2.19))

$$p_{X|Y}(x|y) \stackrel{\text{定義}}{=} \frac{p(x, y)}{p_Y(y)} = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_{x'} p_{Y|X}(y|x')p_X(x')},$$

$$p_{Y|X}(y|x) \stackrel{\text{定義}}{=} \frac{p(x, y)}{p_X(x)} = \frac{p_{X|Y}(x|y)p_Y(y)}{\sum_{y'} p_{X|Y}(x|y')p_Y(y')}.$$

$p_{X|Y}(x|y)$ を $p_{Y|X}(y|x')$ (と $p_X(x')$) で書き表す式,

$p_{Y|X}(y|x)$ を $p_{X|Y}(x|y')$ (と $p_Y(y')$) で書き表す式.

定義の分子分母を, 乗法公式と全確率の公式で書き直しただけ (なので余分の記憶は不要).

L06-Q5

岩薩林 確率・統計 例題 3.3,3.5,3 章練習問題 2

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. 次の通りとする.

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

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

- 1 同時分布を求めよう
- 2 X の周辺分布を求めよう.
- 3 条件付き確率 $p_{Y|X}(20|2), p_{Y|X}(30|3)$ を求めよう.