確率変数の独立, 和, 独立同分布の和

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　L08(2022-05-30 Mon)

最終更新: Time-stamp: "2022-05-30 Mon 11:30 JST hig"

今日の目標

- 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
- 和の確率分布/母平均値/母分散を計算できる
- 独立同分布の和の母平均値/母分散を計算できる

岩薩林 確率・統計 p.113, 定理 5.2

L07-Q1

Quiz 解答: 条件付き分布

1. $p_{\mathrm{Y}|\mathrm{X}}(0|9) = \frac{p(9,0)}{p_{\mathrm{X}}(9)} = \frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}} = \frac{1}{3}.$
2. $p_{\mathrm{X}|\mathrm{Y}}(9|0) = \frac{p(9,0)}{p_{\mathrm{Y}}(0)} = \frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}+\frac{1}{6}} = \frac{1}{4}.$

L07-Q2

Quiz 解答: 同時分布・条件付き分布・周辺分布

$y\backslash x$	2	3	
20	3/12	1/12	4/12
30	4/12	4/12	8/12
	7/12	5/12	1

L07-Q3

Quiz 解答: ベイズの定理

1.

$$p_{\mathrm{Y|X}}(y|1) = \begin{cases} 0.95 & (y = 10) \\ 0.05 & (y = 20) \end{cases}$$

$$p_{\mathrm{Y|X}}(y|2) = \begin{cases} 0.125 & (y = 10) \\ 0.875 & (y = 20) \end{cases}$$

2.

$y \backslash x$	1	2
10	0.19	0.10
20	0.01	0.70

$$\begin{aligned} p_{\mathrm{X|Y}}(1|10) &= \frac{p_{\mathrm{Y|X}}(10|1)p_{\mathrm{X}}(1)}{\sum_x p_{\mathrm{Y|X}}(10|x)p_{\mathrm{X}}(x)} \\ &= \frac{0.95 \times 0.2}{0.95 \times 0.2 + 0.125 \times 0.8} = \frac{19}{29}. \end{aligned}$$

3.

$y\backslash x$	1	2
10	0.76	0.025
20	0.04	0.175

$$
\begin{aligned}
p_{\mathrm{X}|\mathrm{Y}}(2|20) &= \frac{p_{\mathrm{Y}|\mathrm{X}}(20|2)p_{\mathrm{X}}(2)}{\sum_x p_{\mathrm{Y}|\mathrm{X}}(20|x)p_{\mathrm{X}}(x)} \\
&= \frac{0.875 \times 0.2}{0.05 \times 0.8 + 0.875 \times 0.2} = \frac{35}{43}.
\end{aligned}
$$

L07-Q4

Quiz 解答: ベイズ推定

検査 $Y = \begin{cases} 80 & (\text{陽性}) \\ 20 & (\text{陰性}) \end{cases}$, 病状 $X = \begin{cases} 100 & (\text{病気}) \\ 0 & (\text{病気でない}) \end{cases}$ で表す.

$$
p_{\mathrm{X}}(100) = 0.005, p_{\mathrm{Y}|\mathrm{X}}(80|100) = 0.99, p_{\mathrm{Y}|\mathrm{X}}(80|0) = 0.02
$$

同時確率分布

$x\backslash y$	80	20	計
100	真陽性 0.99×0.005	偽陰性 0.01×0.005	0.005
0	偽陽性 0.02×0.995	真陰性 0.98×0.995	0.995

$$
\begin{aligned}
p_{\mathrm{X}|\mathrm{Y}}(100|80) &= \frac{p_{\mathrm{Y}|\mathrm{X}}(80|100) \times p_{\mathrm{Y}}(100)}{p_{\mathrm{Y}|\mathrm{X}}(80|100) \times p_{\mathrm{Y}}(100) + p_{\mathrm{Y}|\mathrm{X}}(80|0) \times p_{\mathrm{Y}}(0)} \\
&= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.02 \times 0.995} = 0.199185.
\end{aligned}
$$

ここまで来たよ

確率変数が「同じ」対「同分布」再訪

2 つの確率変数 X, Y.

X, Y は同じ　定義:$X = Y$, 各試行で同じ値.

$y \backslash x$	0	1
0	$1-p$	0
1	0	p

- X, Y が同じ, ならば, 同分布 (次)
- 例: コイン 1 枚 (表 (1) の確率 $0 < p < 1$) を 1 回投げた結果に, 2 つの変数名 X, Y をつけた

X, Y は同分布　定義: 周辺分布が等しい $p_{\mathrm{X}}(a) = p_{\mathrm{Y}}(a)$.

$y \backslash x$	0	1
0	$1-\frac{3}{2}p$	$\frac{1}{2}p$
1	$\frac{1}{2}p$	$\frac{1}{2}p$

- 「同じ」とはかぎらない
- $\mathrm{E}[g(X)] = \mathrm{E}[g(Y)]$.

X, Y は同分布 2

例: 形の同じ青色コイン X(表 p) と赤色コイン Y(表 q). $p = q$

$y \backslash x$	0	1	
0	$(1-p)(1-q)$	$p(1-q)$	$1-q$
1	$(1-p)q$	pq	q
	$1-p$	p	1

同分布は置いといて, $p \neq q$ でも, この形の同時分布は特別 (カモ).

- X の確率が Y によらない. Y の確率が X によらない.
  - 条件付き分布 $p_{\mathrm{Y|X}}(y, x) = p_{\mathrm{Y}}(y)$. x の条件によらない.
  - 条件付き分布 $p_{\mathrm{X|Y}}(x, y) = p_{\mathrm{X}}(x)$. y の条件によらない.

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 が独立 (independent) とは次が成立すること.

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

X, Y が独立とは

- すべての行 (周辺分布も) がベクトルとして平行. すべての列 (周辺分布も) がベクトルとして平行
- 条件付き確率が周辺分布に等しい $p_{\mathrm{X}|\mathrm{Y}}(x|\text{何でも}) = p_{\mathrm{X}}(x)$, $p_{\mathrm{Y}|\mathrm{X}}(y|\text{何でも}) = p_{\mathrm{Y}}(y)$.
- ベイズ推定は役立たない. 事前確率=事後確率.
- X, Y が互いに「無関係」であること

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.

独立である例

例

$y\backslash x$	0	1
10	4/9	2/9
11	2/9	1/9

例

赤白のサイコロを各 1 個振って, $X=$ 赤の目, $Y=$ 白の目.

例

下のカードから引いて, $X=$ 色, $Y=$ 数.

$\heartsuit 8$　$\diamondsuit 8$　$\diamondsuit 9$

$\spadesuit 8$　$\clubsuit 8$　$\clubsuit 9$

独立でない例

例 1 さいころを 1 個振って,

$$X = \begin{cases} 0 & (\text{目は 4 以下}) \\ 1 & (\text{他}) \end{cases}, \quad Y = \begin{cases} 0 & (\text{他}) \\ 1 & (\text{目は 5 または 6}) \end{cases}.$$

$X = Y$. 確率変数 X, Y は同じ.

$y \backslash x$	0	1
0	2/3	0
1	0	1/3

例 2 例 1 の X, $Y = -X + 1$.「Y は X の関数」

$y \backslash x$	0	1
0	0	1/3
1	2/3	0

X, Y が独立, はカモ

$x = 1, 2, \cdots, 100, y = 1, 2, \cdots, 200$ 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 周辺の $100 + 200$ 個の数だけおぼえればいい.

次の公式を全部使える

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

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

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)

一部の証明連続型で書くけど, $\int \to \sum$ で離散型でも同様.

$$
\begin{aligned}
&\mathrm{E}[g_1(X)\cdot g_2(Y)]\\
&=\iint g_1(x)g_2(y)\cdot f(x,y)\mathrm{d}y\,\mathrm{d}x\\
&\overset{\text{独立}}{=}\iint g_1(x)g_2(y)\cdot f_{\mathrm{X}}(x)\times f_{\mathrm{Y}}(y)\mathrm{d}y\,\mathrm{d}x\\
&=\int g_1(x)f_{\mathrm{X}}(x)\left(\int g_2(y)\cdot f_{\mathrm{Y}}(y)\mathrm{d}y\right)\mathrm{d}x\\
&=\mathrm{E}[g_1(X)]\times \mathrm{E}[g_2(Y)]
\end{aligned}
$$

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

L08-Q1

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)

ここまで来たよ

2 次元離散型確率変数の和

$X = X_1, Y = X_2$: 確率変数.
$S = X_1 + X_2$ とする.
$\mathrm{E}[S] = \mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2]$.
$\mathrm{V}[S] = \mathrm{V}[X_1 + X_2] = \mathrm{V}[X_1] + \mathrm{V}[X_2] + 2\mathrm{Cov}[X_1, X_2] \overset{独立}{=} \mathrm{V}[X_1] + \mathrm{V}[X_2]$.

Quiz(2 次元の離散型確率変数の同時分布と周辺分布)

離散型確率変数 X_1, X_2 の同時分布 $p(x_1, x_2)$ が次の表で与えられる

$x_1 \backslash x_2$	1	2	3
4	4/32	8/32	16/32
5	3/32	1/32	0

確率変数 $S = X_1 + X_2$ の確率関数 $p_S(s)$ を求めよう.

$S = X_1 + X_2$ の確率関数 $p_{\mathrm{S}}(s)$ は?
確率 $P(S = s) = p_{\mathrm{S}}(s)$ となるべき.

$$P(S = s) \overset{\text{定義}}{=} \sum_{x_1} \sum_{x_2} \mathrm{I}_{[X_1 + X_2 = s]}(x_1, x_2) p(x_1, x_2)$$

$$= \sum_{x_1 = -\infty}^{+\infty} p(x_1, s - x_1)$$

特徴関数 $\mathrm{I}_{[\text{ナントカ}]}(x)$ は, ナントカが成立するとき 1, しないとき 0.

和の確率関数

離散型確率変数 X_1, X_2 の同時分布 $p(x_1, x_2)$ のとき, 和 $S = X + Y$ の確率関数は,

$$p_{\mathrm{S}}(s) = \sum_{x_1 = -\infty}^{+\infty} p(x_1, s - x_1) \overset{\text{独立}}{=} \sum_{x_1} p_{\mathrm{X}_1}(x_1) p_{\mathrm{X}_2}(s - x_1).$$

最右辺の形 たたみ込み, 合成積 (convolution)

ここまで来たよ

8 条件付き確率とベイズの公式

8 確率変数の独立, 和, 独立同分布の和
- 確率変数の独立性
- 確率変数の和
- 独立同分布にしたがう確率変数の和

独立同分布

$X = X_1, Y = X_2$ は独立同分布 定義: 独立かつ同分布

例: 形の同じ青色コイン X_1 と赤色コイン X_2.
表がでる (1) の確率 p

$x_2 \backslash x_1$	0	1
0	$(1-p)^2$	$p(1-p)$
1	$p(1-p)$	p^2

独立同分布 (i.i.d.) 岩薩林 確率・統計 定理 4.2(p.87) の仮定,p.113

離散型/連続型確率変数 $X_1, X_2, \ldots, X_n$ が, たがいに独立で, すべて同分布に従う (同じ周辺分布 $p_{\mathrm{X_i}}(x_i) = p(x)$ を持つ) とする.
このとき $X_1, \ldots, X_n$ は独立同分布に従う (i.i.d.=independent and identically-distributed) という.

母期待値も共通になる. $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$.
例: 同じ形の n 色のサイコロ. X_1=青色, X_2=緑色, $\cdots$.

L08-Q2

Quiz(独立同分布の母期待値母分散)

確率変数 X_1, X_2, X_3 は独立同分布に従い, $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$ とする.
次の母期待値, 母分散を μ と σ^2 で表そう.

1. $\mathrm{E}[X_1 + X_1 + X_1]$
2. $\mathrm{E}[X_1 + X_2 + X_3]$
3. $\mathrm{E}[X_1 + X_1 + X_2]$
4. $\mathrm{V}[X_1 + X_1 + X_1]$
5. $\mathrm{V}[X_1 + X_2 + X_3]$
6. $\mathrm{V}[X_1 + X_1 + X_2]$
7. $\mathrm{E}[(X_1 + X_2 + X_3)^2]$

独立同分布にしたがう確率変数の和の性質 岩薩林 確率・統計 例題 4.6

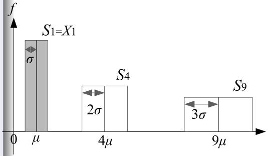
i.i.d にしたがう確率変数の和

$X_1, \ldots, X_n$: i.i.d. 母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$.
和の確率変数 $S_n = X_1 + \cdots + X_n$.

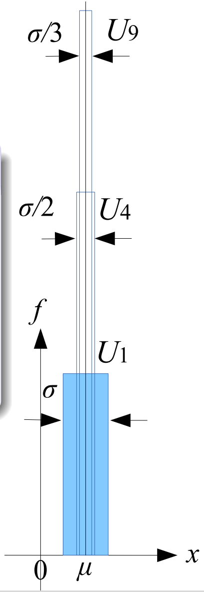
$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同分布}}{=} n \times \mu.$$

$$\mathrm{V}[S_n] \overset{\text{独立}}{=} \sum_{i=1}^{n} \mathrm{V}[X_i] \overset{\text{同分布}}{=} n \times \sigma^2$$

S_n の確率密度関数はこんな感じ?



U_n の確率密度関数はこんな感じ?



i.i.d にしたがう確率変数の和の $1/n$

$X_1, \ldots, X_n$: i.i.d.
新しい確率変数: $U_n = \frac{1}{n} S_n = \frac{1}{n}(X_1 + \cdots + X_n)$

$$\mathrm{E}[U_n] = \mathrm{E}\left[\frac{1}{n} S_n\right] \overset{\text{同分布}}{=} \frac{1}{n} \times n \times \mu.$$

$$\mathrm{V}[U_n] = \mathrm{V}\left[\frac{1}{n} S_n\right] \overset{\text{iid}}{=} \left(\frac{1}{n}\right)^2 \times n \times \sigma^2.$$

L08-Q3

Quiz(独立同分布にしたがう確率変数の和)

ダーツで, 1 回投げるごとに点数が得られ, n 回投げた合計点で競うルールでプレイしている.
あるプレイヤーの i 回目の点数を確率変数 X_i とすると, X_i は独立同分布にしたがい, $\mathrm{E}[X_i] = 80, \mathrm{V}[X_i] = 200$ だという.
このプレイヤーが n 回投げた合計点を確率変数 $S_n = \sum_{i=1}^{n} X_i$ とする.

1. $\mathrm{E}[S_{10}]$, $\mathrm{V}[S_{10}]$ を求めよう.
2. $\mathrm{E}[\frac{1}{10}S_{10}]$, $\mathrm{V}[\frac{1}{10}S_{10}]$ を求めよう.
3. $\mathrm{E}[\frac{1}{100}S_{100}]$, $\mathrm{V}[\frac{1}{100}S_{100}]$ を求めよう.