確率変数の独立性・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　L06(2021-05-19 Wed)

最終更新: Time-stamp: "2021-05-19 Wed 06:48 JST hig"

今日の目標

- ベイズ推定ができる 岩薩林 確率・統計 p.42
- 真/偽 陽/陰性, 混同行列を説明できる 岩薩林 確率・統計 p.42
- 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
- 独立同分布の和の母平均値/母分散を計算できる

L05-Q1
Quiz 解答:条件付き分布

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

L05-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

L05-Q3
Quiz 解答:ベイズの公式

$y\backslash x$	1	2
10	$\frac{1}{3}\cdot\frac{1}{5}$	$\frac{2}{3}\cdot\frac{4}{5}$
20	$\frac{4}{13}\cdot\frac{4}{5}$	$\frac{9}{13}\cdot\frac{4}{5}$

1.
2. $$\begin{aligned} p_{\mathrm{Y|X}}(10|1) &= \frac{p_{\mathrm{X|Y}}(1|10)p_{\mathrm{Y}}(10)}{p_{\mathrm{X|Y}}(1|10)p_{\mathrm{Y}}(10)+p_{\mathrm{X|Y}}(1|20)p_{\mathrm{Y}}(20)} \\ &= \frac{\frac{1}{3}\cdot\frac{1}{5}}{\frac{1}{3}\cdot\frac{1}{5}+\frac{4}{13}\cdot\frac{4}{5}}. \end{aligned}$$
3. $$\begin{aligned} p_{\mathrm{Y|X}}(10|2) &= \frac{p_{\mathrm{X|Y}}(2|10)p_{\mathrm{Y}}(10)}{p_{\mathrm{X|Y}}(2|10)p_{\mathrm{Y}}(10)+p_{\mathrm{X|Y}}(2|20)p_{\mathrm{Y}}(20)} \\ &= \frac{\frac{2}{3}\cdot\frac{4}{5}}{\frac{2}{3}\cdot\frac{4}{5}+\frac{9}{13}\cdot\frac{4}{5}}. \end{aligned}$$

L05-Q4
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

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

3

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

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

ここまで来たよ

L06-Q1

Quiz(ベイズ推定)

袋に何個かの色つきボールが入っている．ボールを割ると，当たり外れの記された紙がでてくる．
当たりのボールのうち赤いボールが $\frac{1}{10}$, 白いボールが $\frac{9}{10}$,
外れのボールのうち赤いボールが $\frac{7}{10}$, 白いボールが $\frac{3}{10}$ とわかっている
袋の中の当たりのボールは, $\frac{2}{10}$ とわかっている．ボールを 1 個取り出したときの当たりの確率 (事前確率) は $\frac{2}{10}$ である．
ボール 1 個を取り出したところ, 赤いボールだった．このとき, 当たりの確率 (事後確率) を求めよう．

岩薩林 確率・統計 例題 2.5, 例題 2.6,2 章練習問題 4,2 章練習問題 5

ベイズ的な考え方

$$\boxed{\text{事後確率 } p_{X|Y}(x|y)} \quad \longleftarrow \quad \boxed{\text{事前確率 } p_X(x)}$$

$$\uparrow$$

$$\boxed{\text{情報 } Y = y}$$

主観確率

事前確率に主観を許す考え方．結果として，事後確率にも主観が含まれる．

L06-Q2

TA Prob and Sol:ベイズ推定

ある病気の人の割合は全体の 0.005 と思われている.
検査では, 病気の人の 0.99 は陽性となり (真陽性), 0.01 の人は陰性になる (偽陰性). また, 病気でない人の 0.02 は (誤って) 陽性となり (偽陽性), 0.98 の人は陰性になる (真陽性).
検査後に陽性となった場合, その人が病気である確率を求めよう.

略解

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

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

混同行列 confision matrix

- 検査の性能を表現する数値は 1 通りではない.
- 現実には本当の確率はわからなくて, 人数の表 (混同行列) で記録する
  - 今後出てくる, 標本抽出, 推定の考えが混ざっている

	検査で陽性	検査で陰性
病気	○ (TP 真陽性)	× (FN 見逃し, 偽陰性)
病気でない	× (FP 誤検出, 偽陽性)	○ (TN 真陰性)

真/偽=True/False, 陽性/陰性=Postive/Negative
どれも大きい方がいいが, 一つを大きくすると他が小さくなりがち

- 感度=再現率=真陽性率=recall=sensitivity= $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
- 特異度=真陰性率=specificity= $\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
- 適合率=precision= $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
- 正解率=精度=accuracy $=\frac{\mathrm{TP}+\mathrm{TN}}{\text{すべて}}$
- F 値=recall と precision の調和平均

病気に限らない. 例:猫画像判定 AI の性能, 陽性=画像は猫, 検査で陽性=AI が画像は猫と判定

ここまで来たよ

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 が同時確率分布 $p(x, y)$ 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.

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

X, Y が独立とは,

X, Y が互いに「無関係」であること

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 1 サイコロを 1 個振って, 目 (1–6) を考える. 岩薩林 確率・統計 例題 3.4(p.57)

$$X = \begin{cases} 0 & (\text{目は 4 以下}) \\ 1 & (\text{目は 5 以上}) \end{cases}, \quad Y = \begin{cases} 10 & (\text{目は偶数}) \\ 11 & (\text{目は奇数}) \end{cases}$$

$y \backslash x$	0	1
10	2/6	1/6
11	2/6	1/6

例 2

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

例 3 赤白のサイコロを 2 個振って,
X = 赤の目, Y = 白の目.

独立とは

同時確率分布が, 周辺分布だけから (積で) 決まっちゃうこと

同時確率分布のすべての行の比が同じ, すべての列の比が同じこと

条件付き確率が周辺確率と等しいこと

$p_{X|Y}(x|\text{何でも}) = p_X(x), \quad p_{Y|X}(y|\text{何でも}) = p_Y(y).$
独立だとベイズ推定は役立たない. 事前確率=事後確率, 無関係.

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.

同分布より厳しい条件

$y\backslash x$	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

例 3 さいころを 1 個振って,

$$X = \begin{cases} 0 & (\text{他}) \\ 1 & (\text{目は 3 の倍数}) \end{cases}, \quad Y = \begin{cases} 0 & (\text{他}) \\ 1 & (\text{目は 5 の倍数}) \end{cases}.$$

$y\backslash x$	0	1
0	3/6	2/6
1	1/6	0

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 周辺の $100 + 200$ 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.

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

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

L06-Q3

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

5 条件付き確率とベイズの定理

6 確率変数の独立性・独立同分布の和
- ベイズ推定
- 確率変数の独立性
- 独立同分布
- 独立同分布にしたがう確率変数の和

確率変数が「同じ」対「同分布」

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.

- X, Y は独立でない
- 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

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

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

- 「同じ」とはかぎらない
- $\mathrm{E}[g(X)] = \mathrm{E}[g(Y)]$.
- 例:コイン 1 枚 X, 表 (1) の確率 $p = 1/2$. $Y = 1 - X$. これは独立でない例.

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

X, Y が同分布であり, (同じでなく) 独立であることはありうる.

$X = X_1, Y = X_2$ 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 と赤色コイン Y.
表 (1) の確率 p

$y \backslash x$	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 (同じ確率関数 $p(x)$ 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.

例: 同じ形の n 色のサイコロ. X_1=青色, X_2=緑色, $\cdots$,

(復習) 確率変数の和の母平均値と母分散

確率変数 $X_1, X_2, Y = aX_1 + bX_2$ を考える.
いつでも $\mathrm{E}[Y] = \mathrm{E}[aX_1 + bX_2] \overset{\text{E5,E3}}{=} a\mathrm{E}[X_1] + b\mathrm{E}[X_2]$.
X_1, X_2 が独立のとき $\mathrm{V}[Y] = \mathrm{V}[aX_1 + bX_2] \overset{\text{IC4}}{=} a^2\mathrm{V}[X_1] + b^2\mathrm{V}[X_2]$.

L06-Q4

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. $\mathrm{E}[(X_1 + X_2 + X_3)^2]$

ここまで来たよ

5 条件付き確率とベイズの定理

6 確率変数の独立性・独立同分布の和
- ベイズ推定
- 確率変数の独立性
- 独立同分布
- 独立同分布にしたがう確率変数の和

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

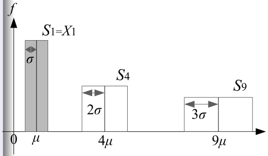
i.i.d にしたがう確率変数の和

$X_1, \ldots, X_n$: 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 $S_n = X_1 + \cdots + X_n$.

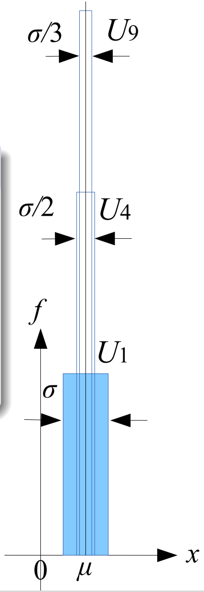
$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同分布}}{=} n \times \mu.$$

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

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: $U_n = \frac{1}{n}S_n = \frac{1}{n}(X_1 + \cdots + X_n)$

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

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

L06-Q5

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.