条件付き確率とベイズの公式

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 L07(2023-05-22 Mon)

最終更新: Time-stamp: "2023-05-22 Mon 11:36 JST hig"

今日の目標

- 条件付き確率 岩薩林 確率・統計 §2.3, p.59, 同時確率分布, 周辺分布の間の計算ができる
- ベイズの定理 岩薩林 確率・統計 p.42 で, 2 つの条件付き確率の間の計算ができる

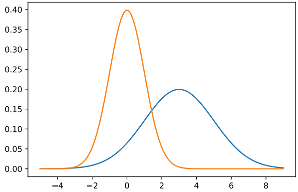
L06-Q1

L06-Q2

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. $P(1 \leq X \leq 7) = P(\frac{1-3}{2} \leq Z \leq \frac{7-3}{2}) = F_{\mathrm{Z}}(2) - F_{\mathrm{Z}}(-1) = 0.81859$.
5. $P(3 \leq X \leq 9) = P(\frac{3-3}{2} \leq Z \leq \frac{9-3}{2})$.
6. $d' = (d-3)/2$. $d' = (F_{\mathrm{Z}}^{-1}(3/4) =$ scipy.stats.norm(loc=0,scale=1).ppf(0.75).

L06-Q3

Quiz 解答: 多変数の確率変数の期待値

1. $p(x,y) =$

$y\backslash x$	1	2	3
0	0	$\frac{2}{12}$	$\frac{1}{12}$
2	$\frac{4}{12}$	0	$\frac{5}{12}$

$g(x,y) = 2x^2 + \mathrm{e}^y =$

$y\backslash x$	1	2	3
0	$2\cdot 1^2+\mathrm{e}^0$	$2\cdot 2^2+\mathrm{e}^0$	$2\cdot 3^2+\mathrm{e}^0$
2	$2\cdot 1^2+\mathrm{e}^2$	$2\cdot 2^2+\mathrm{e}^2$	$2\cdot 3^2+\mathrm{e}^2$

$\mathrm{E}[2X^2+\mathrm{e}^Y] = (2\cdot 1^2+\mathrm{e}^0)0 + (2\cdot 2^2+\mathrm{e}^0)\frac{2}{12} + (2\cdot 3^2+\mathrm{e}^0)\frac{1}{12}$
$+(2\cdot 1^2+\mathrm{e}^2)\frac{4}{12} + (2\cdot 2^2+\mathrm{e}^2)0 + (2\cdot 3^2+\mathrm{e}^2)\frac{5}{12}.$

2. $\frac{4}{12} + 0 + \frac{5}{12} = \frac{9}{12}.$
3.
$$p_{\mathrm{X}}(x) = \begin{cases} 4/12 & (x=1) \\ 2/12 & (x=2) \\ 6/12 & (x=3) \\ 0 & (\text{他}) \end{cases} \qquad p_{\mathrm{Y}}(y) = \begin{cases} 3/12 & (y=0) \\ 9/12 & (y=2) \\ 0 & (\text{他}) \end{cases}$$

4. (1 の別解) $\mathrm{E}[2X^2+\mathrm{e}^Y] = 2\mathrm{E}[X^2] + \mathrm{E}[\mathrm{e}^Y] =$ 周辺分布で計算.

ここまで来たよ

6 一般の正規分布・多次元の確率変数

7 条件付き確率とベイズの公式
- 条件付き確率
- ベイズの定理
- ベイズ推定

同時確率分布と周辺分布 岩薩林 確率・統計 §3.3

X, Y 確率変数
同時確率分布

$p(x,y) = P(X = x$かつ$Y = y) =$

$y\backslash x$	150	160	165
45	3/8	0	1/12
50	1/8	1/3	1/12

コンマ前が X, 後が Y の値. 明示のため $p_{\mathrm{XY}}(x,y)$ と書くことも.
周辺確率分布 $p_{\mathrm{X}}(x) = P(X = x), p_{\mathrm{Y}}(y) = P(Y = y)$
意味 確率の小計, X の値を無視した Y の値だけの分布, その逆
確率変数が等しいとは「$X = 3Y + 15$」のとき,

$p(x,y) =$

$y\backslash x$	150	160	165
45	p	0	0
50	0	0	$1-p$

の形に限られる.
大注意 表の縦横, 変数名 X, Y には意味がなく, 問題などでは入れ替えたりする.

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(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

$P(B|A)$ を P (of)B 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

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

例 $p_{\mathrm{X}|\mathrm{Y}}(150|45), p_{\mathrm{Y}|\mathrm{X}}(45|150)$．縦棒の前だけ見る (縦棒の後ろを固定しする) 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))

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

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

L07-Q1

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, $Y=0(\text{赤札}),1(\text{黒札})$ 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

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)

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

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)

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

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

L07-Q2 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．

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

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

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 $p_{\mathrm{Y}|\mathrm{X}}(20|2), p_{\mathrm{Y}|\mathrm{X}}(30|3)$ を求めよう．

最後の問は次のセクションに延期.

ここまで来たよ

例えばこんな問 I

L07-Q3

Quiz(ベイズの定理)

外見で区別できない, 甘い品種 1 と渋い品種 2 の柿がある.
甘い品種 1 は, 確率 0.95 で赤に, 確率 0.05 で黄色になる.
渋い品種 2 は, 確率 0.125 で赤に, 確率 0.875 で黄色になる.

1. かごの柿の 1/5 が甘い柿である. いま, 無作為に 1 個の柿を取りだしたところ, 赤い柿だった. 取り出した赤い柿が甘い確率を求めよう.

ベイズの定理 I

定理 (ベイズの定理 $\boxed{\text{岩薩林 確率・統計 (2.18),(2.19)}}$)

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

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

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

L07-Q4

ベイズの定理

さっきの問で, 条件付き確率 $p_{\mathrm{Y}|\mathrm{X}}(20|2), p_{\mathrm{Y}|\mathrm{X}}(30|3)$ を求めよう.

ここまで来たよ

L07-Q5

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

ベイズ推定の考え方

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

$\uparrow$

$\boxed{\text{情報 } Y = y}$

追加の情報が得られるたびに，事後確率は正確になっていく．

主観確率事前確率に主観を許す考え方．事後確率にも主観が含まれる．

L07-Q6

Quiz(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).

1.

1 回の検査で陽性となった場合, その人が病気である確率を求めよう.
2 回検査して 2 回とも陽性の場合は?