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(2022-05-23 Mon)

最終更新: Time-stamp: ”2022-05-23 Mon 07:19 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 以降ぜんぶ
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	$\frac{5}{12}$

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

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

$$\mathrm{E}[X^2+\mathrm{e}^Y] = (1^2+\mathrm{e}^0)0 + (2^2+\mathrm{e}^0)\tfrac{2}{12} + (3^2+\mathrm{e}^0)\tfrac{1}{12}$$
$$+(1^2+\mathrm{e}^2)\tfrac{4}{12} + (2^2+\mathrm{e}^2)0 + (3^2+\mathrm{e}^2)\tfrac{5}{12}$$
$$= \tfrac{11}{2} + \tfrac{1}{4} + \mathrm{e}^2\tfrac{3}{4}.$$

2. $g(x,y) = \mathrm{I}_{[XY\geq 2]}(X,Y) =$

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

$$\mathrm{E}[\mathrm{I}_{[XY\geq 2]}(X,Y)] = 0\cdot 0 + 0\cdot\tfrac{2}{12} + 0\cdot\tfrac{1}{12} + 1\cdot\tfrac{4}{12} + 1\cdot 0 + 1\cdot\tfrac{5}{12} = \tfrac{9}{12} = \tfrac{3}{4}.$$

3.

$$p_X(x) = \begin{cases} \frac{1}{3} & (x=1) \\ \frac{1}{6} & (x=2) \\ \frac{1}{2} & (x=3) \end{cases} \qquad p_Y(y) = \begin{cases} \frac{1}{4} & (y=0) \\ \frac{3}{4} & (y=2) \end{cases}$$

4. $\mathrm{E}[X^2 + \mathrm{e}^Y] = \mathrm{E}[X^2] + \mathrm{E}[\mathrm{e}^Y] =$
$(1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{6} + 3^2 \frac{1}{2}) + (\mathrm{e}^0 \cdot \frac{1}{4} + \mathrm{e}^2 \cdot \frac{3}{4}) = \frac{11}{2} + (\frac{1}{4} + \mathrm{e}^2 \frac{3}{4})$.
5. $\mathrm{E}[X^2] = \frac{11}{2}, \mathrm{E}[X] = \frac{13}{6}$. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{29}{36}$.
$\mathrm{E}[Y^2] = 3, \mathrm{E}[Y] = \frac{3}{2}$. $\mathrm{V}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = \frac{3}{4}$.
これらの量は, 同時分布より周辺分布から計算した方が楽.
6. $g(x,y) = xy$ として, 同時分布を使って, $\mathrm{E}[XY] = \frac{19}{6}$.
7. $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \frac{19}{6} - \frac{13}{6} \cdot \frac{3}{2} = -\frac{1}{12}$.
8. $\mathrm{V}[X+Y] = \mathrm{E}[(X+Y)^2] - \mathrm{E}[X+Y]^2 =$
$\mathrm{E}[X^2 + 2XY + Y^2] - (\mathrm{E}[X] + \mathrm{E}[Y])^2 =$
$\mathrm{E}[X^2] + 2\mathrm{E}[XY] + \mathrm{E}[Y^2] - (\mathrm{E}[X]^2 + 2\mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}[Y]^2) = \frac{25}{18}$.
この右辺は, $\mathrm{V}[X] + 2\mathrm{Cov}[X,Y] + \mathrm{V}[Y]$ の形にまとまる.

ここまで来たよ

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

$p(x, y) = P(X = x かつ Y = y) =$

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

コンマの前が X, 後ろが Y の値. はっきりさせるには $p_{\mathrm{XY}}(x, y)$ と書いてもいい.

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

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.

「同じ」確率変数「同じ」確率分布とは?

例で　離散型確率変数 X, Y, 確率関数 $p(x,y) = $

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

周辺分布 $p_X(x) = \begin{cases} \frac{1}{4} & (x=2) \\ \frac{3}{4} & (x=3) \end{cases}, p_Y(y) = \begin{cases} \frac{1}{4} & (y=2) \\ \frac{3}{4} & (y=3) \end{cases}$.

X, Y の周辺分布の形が同じ $\rightsquigarrow$「X, Y は同分布にしたがう」という.
同分布にしたがう $\Rightarrow \mathrm{E}[g(X)] = \mathrm{E}[g(Y)]$.
$\mathrm{E}[X+Y] = \frac{11}{4} + \frac{11}{4} = \mathrm{E}[X+X]$ だけど, これは, 性質
$\mathrm{E}[g_1(X,Y) + g_2(X,Y)] = \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_1(X,Y)]$ による.

同分布にしたがう $\not\Leftrightarrow X = Y \Leftrightarrow$ 同じ確率変数
代入 $Y = X$ はできない: $\mathrm{E}[XY] = \frac{30}{4} \neq \frac{31}{4} = \mathrm{E}[XX]$.
$\mathrm{Cov}[X,X] = \mathrm{V}[X] \geq 0, = \frac{31}{4} - (\frac{11}{4})^2 = \frac{3}{16}$: X のちらばり$=X$: 大で X: 大となる傾向
$\mathrm{Cov}[X,Y] = \frac{30}{4} - (\frac{11}{4})^2 = -\frac{1}{16}$: X: 大で Y: 大となる傾向
$\rightsquigarrow$ 独立性

何回かあと

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「|=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. 全事象に対して 1.

離散型確率変数に対する条件付き確率

同時確率分布

$p(x,y) = P(X = x \text{かつ} Y = y) =$

$y \backslash x$	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 $X = x$ の条件付き確率の記号
$P(X = x|Y = y) = p_{\mathrm{X|Y}}(x \text{ の値}|y \text{ の値})$

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

条件付き確率を同時確率分布と周辺分布から求める

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

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

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

同時確率分布，周辺分布，条件付き確率の関係

同時確率分布との関係 岩薩林 確率・統計 乗法公式 (2.13)

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

さらに周辺分布との関係 岩薩林 確率・統計 全確率の法則 (2.17)

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

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

L07-Q2

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

ここまで来たよ

7 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

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

ベイズの定理 岩薩林 確率・統計 (2.18),(2.19)

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

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

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

7 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

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

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

$$\uparrow$$

$$\boxed{\text{情報 } Y = y}$$

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 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?(2 回の検査は ‘独立’ とする)

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= $\frac{\text{TP}}{\text{TP}+\text{FN}}$
- 特異度=真陰性率=specificity= $\frac{\text{TN}}{\text{TN}+\text{FP}}$
- 適合率=precision= $\frac{\text{TP}}{\text{TP}+\text{FP}}$
- 正解率=精度=accuracy $= \frac{\text{TP}+\text{TN}}{\text{すべて}}$
- 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