独立性の検定・ピボットテーブル

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 II L04(2016-05-12 Thu)

最終更新: Time-stamp: "2016-05-12 Thu 12:48 JST hig"

今日の目標

- カテゴリ 2 変数の標本から, ピアソンの χ^2, クラメールの連関係数 V が計算できる
- 独立性のカイ二乗検定ができる
- Excel でピボットテーブルを使ってクロス集計表を作れる

http://hig3.net

L03-Q1
Quiz 解答:ベイズの公式

1.

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

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

2.

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

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

3.

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

L03-Q2

Quiz 解答:ベイズ推定

Y を色, X を当落とすると,

$$
\begin{aligned}
&P(X = \text{落} \,|Y = \text{赤}) \\
=&\frac{P(Y = \text{赤} \,|X = \text{当})P(X = \text{当})}{P(Y = \text{赤} \,|X = \text{当})P(X = \text{当}) + P(Y = \text{赤} \,|X = \text{落})P(X = \text{落})} \\
=&\frac{\frac{7}{10}\frac{8}{10}}{\frac{1}{10}\frac{2}{10}+\frac{7}{10}\frac{8}{10}} = \frac{28}{29}.
\end{aligned}
$$

$Y\backslash X$	当	落
赤	$\frac{1}{10}\cdot\frac{2}{10}$	$\frac{7}{10}\cdot\frac{8}{10}$
白	$\frac{9}{10}\cdot\frac{2}{10}$	$\frac{3}{10}\cdot\frac{8}{10}$
合計	$\frac{2}{10}$	$\frac{8}{10}$

L03-Q3

Quiz 解答:ピアソンの χ^2 と適合度の検定

1. $$\chi^2 = \frac{(24 \times \frac{6}{12} - 8)^2}{24 \times \frac{6}{12}} + \cdots = \frac{16}{3}.$$

2. 自由度は $k = C - 1 = 4 - 1$. 有意水準 $\alpha = 0.05$ で, $\chi_\alpha(4-1) = 7.815 > \frac{16}{3}$. よって, 適合するという帰無仮説は棄却できない.

L03-Q4

Quiz 解答:ピアソンの χ^2 と適合度の検定

1. $$\chi^2 = \frac{(14 - 60 \cdot \frac{1}{6})^2}{60 \cdot \frac{1}{6}} + \cdots = \frac{42}{10}$$

   .

2. 有意水準 $\alpha = 0.05$ で
適合度のカイ二乗検定を行う
帰無仮説を, 標本は確率各面 $\frac{1}{6}$ のサイコロで抽出された, とする.
帰無仮説のもとで, ピアソンの適合度基準は χ^2 は自由度
$k = C - 1 = 6 - 1$ のカイ二乗分布に従う. これを検定統計量として
用いる.
標本に対して, 上の通り, $\chi^2 = \frac{42}{10}$ である.
カイ二乗分布表を見ると, $\chi_{0.05}(6-1) = 11.07 > 4.2$ なので, 帰無仮
説は棄却できない.

ここまで来たよ

1. 適合度の検定

2. 独立性の検定・ピボットテーブル
   - 質的変数が 2 つ:独立性の指標
   - 独立性の検定
   - クラメールの連関係数 V

2 つのカテゴリカル変数

未知の母分布

$Y \setminus X$	A 型	A 型以外
女子	P(血液型=A 型, 性別=女)	P(血液型=A 型以外, 性別=女)
男子	P(血液型=A 型, 性別=男)	P(血液型=A 型以外, 性別=女)

標本

出席番号	血液型	性別
1	A 型以外	男
2	A 型以外	女
⋮	⋮	⋮
12	A 型	女

標本サイズ $N = 12$

分割表, クロス集計表

Excel

	A 型	A 型以外
女子	$n_{11} = 1$	$n_{12} = 2$
男子	$n_{21} = 4$	$n_{22} = 5$

度数 n_{ij}, $1 \leq i \leq c, 1 \leq j \leq r$. 行数 r, 列数 c.

性別と血液型は関係ある?

'関係ある' 度を考えたい. 将来的には検定に使いたい.
関係ある の否定は,

- 関係ない
- 性別と血液型は確率変数として独立である
-

$$P(\text{血液型}{=}\text{A 型}, \text{性別}{=}\text{男}) = P(\text{血液型}{=}\text{A 型}) \times P(\text{性別}{=}\text{男})$$
$$f_{\mathrm{XY}}(x, y) = f_{\mathrm{X}}(x) \times f_{\mathrm{Y}}(y).$$

標本の周辺分布

母分布の周辺分布を，標本の周辺分布で推定

$y \setminus x$	A 型	A 型以外	計
女子	1	2	3
男子	4	5	9
計	5	7	12

- P(性別=女) は $p_1 = \frac{3}{12}$ くらい
- P(血液型=A 型) は $q_1 = \frac{5}{12}$ くらい

期待度数

もし，性別と血液型が無関係 (=独立) なら．A 型の女子は

$$\text{期待度数} = N \times p_1 \times q_1 = 12 \times \frac{3}{12} \times \frac{5}{12} = 1.25$$

人くらいのはず

「独立でない度」:ピアソンの χ^2

期待度数

	A 型	A 型以外	計
女子	Np_1q_1	Np_1q_2	Np_1
男子	Np_2q_1	Np_2q_2	Np_2
計	Nq_1	Nq_2	N

(ずれ)^2 $= \sum$(度数 − 期待度数)^2

「独立でない度」:ピアソンの χ^2(カイ二乗)

$p_i\ (i=1,\ldots,r)$, $q_j\ (j=1,\ldots,c)$: 標本から推定した周辺分布.

$$\chi^2 = \frac{(\text{度数} - \text{期待度数})^2}{\text{期待度数}}\text{の合計} = \sum_{1\le i\le r, 1\le j\le c} \frac{(n_{ij} - Np_iq_j)^2}{Np_iq_j}$$

適合度基準との関係

いまの場合

$$\chi^2 = \frac{(1-1.25)^2}{1.25} + \frac{(2-1.75)^2}{1.75} + \frac{(4-3.75)^2}{3.75} + \frac{(5-5.25)^2}{5.25} = 0.11685.$$

ピアソンの χ^2(カイ二乗) の性質

- $0 \leq \chi^2$.
- 大きいほど '独立でなさそう'
- 実は, 自由度 $(r-1)(c-1)$ のカイ二乗分布にしたがう.

Example

Excel で分割表を作って χ^2 を求めようピボットテーブルという Excel の機能を使うのが便利
RaMMoodle `https://el.math.ryukoku.ac.jp/moodle` のデータをクロス集計表にして, 独立性の検定をして, 課題にアップロード.
標本のデータ部分を選択して, 挿入 > ピボットテーブル.

ここまで来たよ

1. 適合度の検定

2. 独立性の検定・ピボットテーブル
   - 質的変数が 2 つ:独立性の指標
   - 独立性の検定
   - クラメールの連関係数 V

独立性の検定

1. 「有意水準 $\alpha = \ldots$ で」，
2. 「独立性のカイ二乗検定を行う」
3. 「帰無仮説を，‘X,Y が独立な母集団から抽出された’ とする」
4. 「帰無仮説の本で検定統計量 ピアソンの χ^2 は自由度 $(c-1)(r-1)$ のカイ二乗分布にしたがう．これを検定統計量として用いる」
5. 「標本に対して $\chi^2 = \ldots$ である」
6. 「χ^2 より極端な値になる確率 p は，カイ二乗分布表より，α 以上/未満なので帰無仮説を棄却する/しない (X と Y には関係がある/あるとは言えない)

L04-Q1

Quiz(ピアソンの χ^2 と独立性の検定)

日本人の高校生から標本を抽出し, 6 人を, 右利きかどうか, 早生まれかどうかで分類すると, 度数 (人数) は下の表のようになった.

	右利き	右利きでない
早生まれ	1	1
早生まれでない	3	1

1. ピアソンの χ^2 を求めよう.
2. 早生まれかどうかと右利きであるかどうかは独立か. 有意水準 $\alpha = 0.05$ で, 独立性のカイ二乗検定を行って判定しよう. 「○○○(不等式) なので, 帰無仮説を棄却する/しない. X と Y には関係がある/あるとは言えない」の形で答えよう.

ここまで来たよ

クラメールの連関係数 V

クラメールの連関係数 V

χ^2:ピアソンの χ^2, N:サンプルサイズ.

$$V = \sqrt{\frac{\chi^2}{N}}$$

例 $V = \sqrt{\frac{0.11685}{12}} = 0.0987$

クラメールの連関係数 V の性質

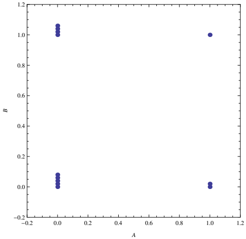
χ^2 を, 相関係数 r みたいに $0 \leq V \leq 1$ を満たすように変換したもの

- $V = 0$ 関係なし
- $V = 1$ 関係あり

相関係数との関係:ダミー変数

- 女子 $A=1$, 男子 $A=0$.
- A 型 $B=1$, A 型以外 $B=0$.

というように量的変数にしちゃえば? …ダミー変数



	A 型	A 型以外
女子	1	2
男子	4	5

$\rightsquigarrow$ 相関係数 r が求まる. 意味あるの?

- 0 と 100 じゃいけないの?
- 0 と 1 を逆にしたら?

2×2 のときの r と連関係数 V の関係

$$|r| = V$$

お知らせ

- 確率統計☆演習 I と同じセッティングで予習問題をやりましょう.
  http://hig3.net → RaMMoodle
  https://el.math.ryukoku.ac.jp/moodle/ → 確率統計☆演習 II(2016)
- チューター/Math ラウンジ 月火水木昼 1-614

https://manaba.ryukoku.ac.jp
マイページの下の方に
manaba 出席カード提出

カイ二乗分布表

有意水準 α, 自由度 k に対して, $\alpha = P(Y > \chi^2_\alpha(k))$ となる $\chi^2_\alpha(k)$ の値の表.

$k\backslash\alpha$	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.00003927	0.0001571	0.0009821	0.003932	0.01579	2.706	3.841	5.024	6.635	7.879
2	0.01003	0.02010	0.05064	0.1026	0.2107	4.605	5.991	7.378	9.210	10.60
3	0.07172	0.1148	0.2158	0.3518	0.5844	6.251	7.815	9.348	11.34	12.84
4	0.2070	0.2971	0.4844	0.7107	1.064	7.779	9.488	11.14	13.28	14.86
5	0.4117	0.5543	0.8312	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	0.6757	0.8721	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	0.9893	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	23.54	26.30	28.85	32.00	34.27
17	5.697	6.408	7.564	8.672	10.09	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.81	37.16
19	6.844	7.633	8.907	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.434	8.260	9.591	10.85	12.44	28.41	31.41	34.17	37.57	40.00
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2