条件付き分布，ベイズ推定

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 L15(2025-07-28 Mon)

最終更新: Time-stamp: "2025-07-28 Mon 07:51 JST hig"

今日の目標

- 第 1 種, 第 2 種の過誤, 有意水準と検出力を説明できる 久保川 統計学入門 §9.3
- 適合度のカイ二乗検定ができる 久保川 統計学入門 §11.1
- ベイズ推定できる 久保川 統計学入門 §3.4

L14-Q1(上側片側)
解答は 確率統計 I(2025)L14 に掲載
L14-Q2 (下側片側)

Quiz 解答: 母比率の片側検定

1. 有意水準 $\alpha = 0.01$ で正規近似による母比率の片側 (二項) 検定を行う.
2. 帰無仮説を母比率 $r = \frac{1}{10}$, 対立仮説を母比率 $r < \frac{1}{10}$ とする.
3. 帰無仮説のもとで, 100 本中のあたりの回数 $T \sim \mathrm{B}(\frac{1}{10}, 100)$. あたりの標本比率を $\hat{r}$ とすると, $Z = \frac{\hat{r}-\frac{1}{10}}{\sqrt{\frac{1}{100}\cdot\frac{1}{10}(1-\frac{1}{10})}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう.
4. この標本に対して $\hat{r} = \frac{5}{100}$ なので, 検定統計量の実現値は $z = -\frac{5}{3} = -1.667$.

5. 棄却限界は $-z_{0.01} = \Phi^{-1}(0.01) =$ scipy.stats.norm(loc=0,scale=1).ppf$(0.01) = -2.326$. 棄却域は $z < -z_{0.01}$.
6. 検定統計量の実現値 z は棄却域に含まれない．帰無仮説を受容する．母比率 $r < \frac{1}{10}$ と結論できない．

p 値による場合

5. p 値は $p = P(Z < -\frac{5}{3}) = \Phi(-\frac{5}{3}) - \Phi(-\infty) =$ scipy.stats.norm(loc=0,scale=1).cdf$(-\frac{5}{3}) = 0.0475$.
6. $0.0475 = p > \alpha = 0.01$ なので帰無仮説を受容する．よって母比率 $r < \frac{1}{10}$ と結論できない．

棄却域に含まれる場合．p 値が α より小さい場合．

6. 帰無仮説を棄却する．母比率 $r < \frac{1}{10}$ と結論する．

ここまで来たよ

母平均値・母分散・母比率・適合度・独立性の検定

	両側	片側
母平均値	**両側 t 検定** 久保川 統計学入門 公式 9.4	**片側 t 検定** 久保川 統計学入門 公式 9.4
母分散	**両側カイ二乗検定** 久保川 統計学入門 公式 9.7	**片側カイ二乗検定** 久保川 統計学入門 公式 9.7
母比率	**両側二項検定** 久保川 統計学入門 (9.7)	**片側二項検定** 久保川 統計学入門 例題 9.10
適合度	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.2
独立性	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.4
母平均値の差	**両側 t 検定** 対応なし (等分散, Welch) 久保川 統計学入門 公式 10.7, あり 久保川 統計学入門 公式 10.5	**片側 t 検定** 対応なし (等分散, Welch), あり
母分散の比	**両側 F 検定** 久保川 統計学入門 公式 10.3	**片側 F 検定**

仮説検定での第１種・第２種の過誤と混同行列 久保川 統計学入門 §9.3

	帰無仮説を棄却, 有意である	帰無仮説を受容, 有意でない
対立仮説が成立 $\mu \neq \mu_0$	確率 $1-\beta$	第２種の過誤 確率 β
帰無仮説が成立 $\mu = \mu_0$	第１種の過誤 確率 α	確率 $1-\alpha$

α, β はどちらも小さくしたいが, 両立しない.
α はほぼゼロ ($\alpha = 0.01$ or 0.05) に固定して, その上で β をなるべく小さくするように試みる習慣.
検定とは, 発色 (検出) したら (まれな誤検出を除いて) ほぼ確かなだけど, 見逃すことはある異常検査薬のようなもの.
α 誤検出率, 有意水準, 危険率
β: 見逃し率 (検出力:$1-\beta$)
類似: 混同行列

カテゴリ変数 久保川 統計学入門 §1.5

量的変数の確率変数 ... 平均や分散に意味がある

- 離散型 確率関数=表 二項分布, ポアソン分布, $\cdots$, x は整数
- 連続型 確率密度関数 正規分布, カイ二乗分布, 指数分布，一様分布, $\cdots$, x は実数

質的変数の確率変数 ... 平均や分散が考えられない
その中でも, 名義変数=カテゴリ (カル) 変数
順序や距離がなくぜんぶが対等. 例: 血液型, 性別, 携帯電話番号, チーム A 型, B 型などがカテゴリ

2 カテゴリなら, 0,1 のように番号を振って離散型と思える
3 カテゴリ以上なら, 順序や間隔によるので離散型には帰着できない.
あえてやるなら one-hot vector.

カテゴリ分布と多項分布 (multinomial distribution)

久保川 統計学入門 §14.1.3

定義 (k カテゴリのカテゴリ分布 $\mathrm{C}(p_1,\dots,p_k)$)

カテゴリ番号 $i=1,\dots,k$ と，パラメタ $p_1+\cdots+p_k=1$. 略記 $=(p_1,\dots,p_k)$ に対して，

$$\text{確率関数} p(i)=P(X=i)=p_i=\begin{cases} p_1 & (i=1) \\ p_2 & (i=2) \\ \vdots & \\ p_k & (i=k) \end{cases}$$

ベルヌイ分布 $Ber(p_1)$ は $k=2$ の $\mathrm{C}(1-p,p)$.

定義 (k カテゴリの多項分布 $Mult_k(n,p_1,\dots,p_k)$)

確率変数は整数値ベクトル $(X_1,\dots,X_k)$. パラメタ $p_1,\dots,p_k,n$.

$$\text{確率関数} P(X_1=x_1,\dots,X_k=x_k)=f(x_1,\dots,x_k)=\begin{cases} \frac{n!}{x_1!\dots x_k!}p_1^{x_1}\dots p_k^{x_k} & (x_i\ge 0, x_1+\cdots+x_k=n) \\ 0 & (\text{他}) \end{cases}$$

カテゴリ分布にしたがう確率変数の n 個の実現値を得て，カテゴリ $1,\dots,k$ をそれぞれ $x_1,\dots,x_k$ 回ずつ得る確率.

	$n=1$	n	ベイズ共役分布	
$k=1$	ベルヌイ分布 $Ber(p)$	二項分布 $Bin(n,p)$	二項検定	ベータ分布
k	カテゴリ分布 $\mathrm{C}(p_1,\dots,p_k)$	多項分布 $Mult_k(1,p_1,\dots,p_k)$	適合度の検定	ディリクレ分布

適合度基準 久保川 統計学入門 §11.1

カテゴリの個数 $k=4$ の例

カテゴリ	O 型	A 型	AB 型	B 型
確率 p_i	$p_1=0.12$	$p_2=0.51$	$p_3=0.17$	$p_4=0.20$

$$\sum_{i=1}^{k} p_i = 1.$$

カテゴリ分布のサイズ $n=12$ の標本

番号	血液型
1	B 型
2	O 型
⋮	⋮
12	A 型

$\to$ 度数分布表は多項分布 $\mathrm{M}(,12)$ の実現値のはず

カテゴリ	O 型	A 型	AB 型	B 型
度数 x_i	$x_1=2$	$x_2=3$	$x_3=6$	$x_4=1$

$$\sum_{i=1}^{k} x_i = n = 12.$$

適合度を表す量

カテゴリ i の期待度数=標本サイズ × 母分布の確率=np_i

命題 (ピアソンの適合度基準 Q 久保川 統計学入門 §11.1)

カテゴリ k 個, カテゴリ分布の確率 $p_i (i = 1, \dots, k)$
サイズ n の標本の度数 $x_i (i = 1, \dots, k)$
サイズ n の標本の, 母分布から考えた期待度数 $np_i (i = 1, \dots, k)$

$$\text{ピアソンの適合度基準} Q(x) = \frac{(\text{度数} - \text{期待度数})^2}{\text{期待度数}} \text{の合計} = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i}$$

が小さいほど, 標本は母分布から予想されるものに近い

命題 (久保川 統計学入門 公式 11.2)

n 大のとき，ピアソンの適合度基準 Q は，自由度 $k-1$ のカイ二乗分布にしたがう.

このことを利用して，帰無仮説「$p_1, \dots, p_k$ のカテゴリ分布にしたがう」対立仮説「したがわない ($\Leftrightarrow p_i$ が 1 個は異なる」で検定が行える 適合度のカイ二乗検定 久保川 統計学入門 §11.1

ここまで来たよ

14 片側検定

15 条件付き分布，ベイズ推定

- 検定の第 1 種・第 2 種の過誤・適合度のカイ二乗検定
- 条件付き確率

復習：同時分布と周辺分布

久保川 統計学入門 §6.1　確率統計 I(2025)L06

X, Y 確率変数

同時分布

$p(x, y) = P(X = x かつ Y = y) =$

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

周辺分布

X, Y の周辺分布 $p_{\mathrm{X}}(x)$,$p_{\mathrm{Y}}(y)$ は，

$$離散型 p_{\mathrm{X}}(x) = p(x, \bullet) = \sum_y p(x, y), \quad p_{\mathrm{Y}}(y) = p(\bullet, y) = \sum_x p(x, y)$$

要するに一方を無視した分布．小計．

条件付き確率 久保川 統計学入門 §3.3

定義 (条件付き確率 久保川 統計学入門 定義 3.7, 公式 3.8)

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 久保川 統計学入門 §6.3

同時確率分布

$p(x,y) = P(X = x \text{かつ} Y = y) =$

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

定義 (離散型確率変数に対する条件付き確率)

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

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 久保川 統計学入門 定義 6.9)

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

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

L15-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.

$\heartsuit$7　$\heartsuit$8　$\heartsuit$9　$\diamond$8　$\spadesuit$9　$\clubsuit$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) 久保川 統計学入門 公式 3.8

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

全確率の法則 (周辺分布との関係) 久保川 統計学入門 §3.4

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

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

ベイズの定理 久保川 統計学入門 §3.4 I

定理 (ベイズの定理 久保川 統計学入門 公式 3.15)

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

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

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

L15-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．

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

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

1. 同時分布を求めよう
2. X の周辺分布を求めよう．
3. 条件付き確率 $p_{\mathrm{Y|X}}(20|2), p_{\mathrm{Y|X}}(30|3)$ を求めよう．

ベイズ推定

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

$\uparrow$

$\boxed{\text{情報 } Y = y}$

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．

例えばこんな問 I

L15-Q3

Quiz(ベイズの推定)

甘い品種 1 と渋い品種 2 の柿が袋に入っている．
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. 袋の柿の 1/5 が甘い品種 1 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品種 1 である確率を求めよう．
2. 袋から取り出すの柿の 1/5 が渋い品種 2 であると考えている．いま，1 個の柿を取りだしたところ，黄色い柿だった．この柿が渋い確率を求めよう．

L15-Q4

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 (真陰性).

1. 1 回の検査で陽性となった人が，病気である確率を求めよう．
2. 1 回の検査で陽性となった人が，さらにもう 1 回検査してやはり陽性だった．その人が病気である確率を求めよう．

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

ふつうじゃない trialL15(締切 2025-08-04 月 20:00)

- trialL15-1 LearnMoodle 練習問題的なものだが，受験は 2 回まで．チェックは使えない
  - ほぼ同内容の予行演習用の練習問題 L15 に合格すると開放されます．
  - 趣旨: 対面でやる機会がない
- trialL15-2 LearnMoodle 作文
  - まずアンケートにご協力お願いします．回答すると開放されます．
  - 学習履歴分析への同意をお願いします

定期試験期間中の任意参加授業内小テスト

用途:LearnMoodle 成績評価方法参照．
2025-08-04 月 09:15-10:15, 参照なし．7-002．事前の出欠連絡不要．理由を問わず, 欠席したときの追試や配慮はなし．Trial L01-08 の出題計画をあわせたもの（注：2024 年度と L08 の内容が異なる）．要学生証, 個人別座席指定．
すべて持込不可．
災害や交通機関などにより全学的に定期試験が中止・延期になるときは同様に中止・延期．

成績評価

考慮対象になるかもしれないものは LearnMoodle 欠席届に 2025-07-30 水までに提出．

メッセージ

- 教科書は売らないで．確率統計 II,III（3 年後期）他で使います．
- 数理と社会（3 年 1Q）でお会いしましょう．