数理統計 補助資料
～ベイズ統計学とは？～

2023年度2学期：　月曜1限, 水曜3限
担当教員：　石垣　司

ベイズ統計学とは？

ベイズの定理を積極的に利用する統計学
- 人間の主観や信念などを情報創出のために利用できる

この授業で紹介するベイズ統計学の有用性
1. 統計的意思決定への応用
2. 情報の逐次更新
3. 柔軟な統計モデル

どれもが深奥な学術的意義と応用上の有用性をもつ。
本授業ではベイズ統計学の核となる部分を紹介

#1　非ベイズ統計学を頻度論や頻度主義的統計学とよぶ。信頼区間、統計的検定などは頻度論的な方法論
#2　頻度主義的統計学とベイズ統計学は哲学的なアプローチが異なる(端的には"確率"や"科学的"の考え方が違う)ため、長い間論争が絶えなかった。ただ、最近ではベイズ統計学の有用性が周知してきたため論争は落ち着いている。どちらが正しい・優れているという話ではない。考え方、有用な場面が異なる。

補足：ベイズ統計学に関する本授業の内容

ベイズ統計学の核となる部分と有用性の紹介
- ベイズ推定, 特に事後分布の推定に関する数理的な内容は学部2年生の学習内容を大きく超えてしまうため詳細の説明は行うことができない。興味のある学生はガイダンスで紹介した参考書を参考にしてほしい。
- とはいえ, How to 本ではないので, 本質的な部分の説明は数学も含めてしっかり紹介する。

(この授業で扱う)統計的な因果関係 #1

統計的因果関係(この授業で扱う因果関係)の表現
- ある原因に応じて結果が確率的に出現する因果関係
- 原因と結果の事象
  結果となる事象の集合 $A = \{a_1, \dots, a_{|A|}\}$
  結果の原因となる事象の集合 $B = \{b_1, \dots, b_{|B|}\}$
  $a_1, \dots, a_{|A|}$ は互いに排反。$b_1, \dots, b_{|B|}$ は互いに排反

【原因】
・インフルエンザワクチン接種(する, しない)
・居住地300Km以内に低気圧(あり, なし)
・意中の人へ告白(する, しない)

【結果～不確実性をもつ】
⇒ インフルエンザ発症 or 非発症
⇒ 頭痛が起こる or 起こらない
⇒ 恋人になる or ならない(保留等含む)

#1　因果関係は必然関係(BならばAでなければならない)ではないので注意
#2　因果関係の定義、統計的因果推定、因果効果は本セメスターの終わりころの回で取り扱う予定

(この授業で扱う)統計的な因果関係 #2

原因に対する結果の出やすさを確率で表現する

- **例：病気と検査結果**
  - $A = \{a_1, a_2\}$（a_1:**検査で陽性**, a_2:**検査で陰性**）
  - $B = \{b_1, b_2\}$（b_1:**罹患している**, b_2:**罹患していない**）
  - $\Pr(B = b_1)$：**ある人が罹患している確率** or **集団の中で罹患している人の割合など, 分析の文脈で使い分ける**
  - $\Pr(A = a_1 | B = b_1)$：**罹患している人が検査で陽性になる確率**

$\Pr(A|B)$

	$B = b_1$	$B = b_2$
$A = a_1$	0.9	0.2
$A = a_2$	0.1	0.8

$\Pr(B)$

$B = b_1$	0.4
$B = b_2$	0.6

重要な注意事項

- **数式や確率表自体は原因と結果を区別&規定できない**
- **確率や条件付確率の意味は内容に合わせて分析者が解釈**
  「ワクチンを打ったけどインフル発症」or「ワクチンを打ったからインフル発症」

ベイズの定理

Thomas Bayes 1702-1761

ベイズの定理

- **事後確率(逆確率)を求める公式**

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)} = \frac{\Pr(A|B)\Pr(B)}{\sum_{i=1}^{|B|}\Pr(A|B = b_i)\Pr(B = b_i)}$$

check!

- $\Pr(B|A)$：**事後確率**, $\Pr(B)$：**事前確率**

事後確率 $\Pr(b_1|a_1)$ の解釈（**今後**,$\Pr(B = b_1|A = a_1)$ を$\Pr(b_1|a_1)$ **と記述**）

- **ある結果 a_1 が生じる原因として互いに排反な b_1 or b_2 のみが考えられる場合、結果 a_1 が生じた原因が b_1 である確率**
  $\Pr$(**原因**), $\Pr$(**結果**|**原因**)**が既知** ⇒ $\Pr$(**原因**|**結果**)**を計算可能**
- $\Pr(b_1|a_1)$：**検査で陽性だった人が罹患している確率**

ベイズの定理を用いた例題 #1

例題：ある検査薬を用いた病気の検査を考える

- **罹患者が陽性反応を示す確率** 99.9 %
- **非罹患者が陽性反応を示す確率** 0.2 %
- **その病気の日本人の罹患率は**100,000**人に**1**人**
- **日本人全体から無作為抽出したある人にその検査薬を用いたところ陽性反応が出た。その人が罹患している確率は？**

- **原因：** b_1:**罹患**, b_2:**非罹患**
- **結果：** a_1:**陽性反応**, a_2:**陰性反応**
- **事前確率：** $\Pr(b_1) = 0.00001$
- **その他の確率：** $\Pr(b_2) = 0.99999$, $\Pr(a_1|b_1) = 0.999$, $\Pr(a_1|b_2) = 0.002$

ベイズの定理を用いた例題 #2

例題の解答（求めるべきは事後確率$\Pr(b_1|a_1)$）

$$\Pr(b_1|a_1) = \frac{\Pr(a_1|b_1)\Pr(b_1)}{\Pr(a_1|b_1)P(b_1) + \Pr(a_1|b_2)\Pr(b_2)}$$
$$= \frac{0.999 \times 0.00001}{0.999 \times 0.00001 + 0.002 \times 0.99999} = \frac{999}{200997} \approx 0.005$$

- **陽性反応の出た人が罹患している確率は** 0.5 %

母集団：日本人全体

罹患者：0.001%
陽性反応率：99.9%

非罹患者：99.999%
陽性反応率：0.2%

陽性反応が出た人の罹患率 =

#メモ　実用上は,「日本人全体から無作為抽出」と「症状が出たので病院に来た人」では統計的な問題も医療上の意義も異なるので注意

ベイズの定理による情報創出の構造

例題の解答から得られる示唆

- 事前確率の値が事後確率の値に影響を与える
- 事後確率がもつ情報は増えている

  検査を受けていない人の罹患率: 0.0001 %

  陽性反応が出た人の罹患率: 0.5 % ⇒ 危険度は検査前の5,000倍

ベイズの定理による情報創出の構造

事後の情報 ⇐ データが持つ情報 ＋ 事前の情報

- ベイズの定理は確率の公理を満たす単なる数式であるが、この構造を自然に表現している

以降は, この情報創出の構造の統計モデルでの利用を考える

ベイズ統計学と統計モデル

統計モデル(復習)

- 関数 $f(y;\boldsymbol{\theta})$ と確率的構造を組み合わせる
- 関数の形はパラメータ $\boldsymbol{\theta}$ の値で決まる
- 統計モデルの推定: データを用いて $\boldsymbol{\theta}$ の値を推定

ベイズ統計学での統計モデル

- パラメータ $\boldsymbol{\theta}$ を確率変数とみなして事前情報をモデル化する
- 事前情報＋データを用いて $\boldsymbol{\theta}$ の値を推定

例: 単回帰モデル　$y = a + bx$

- もしも, パラメータ a と b について何らかの情報を持っているとき, a と b を確率変数とみなして, 確率密度関数でその情報を表現

数学的準備 #1 (復習)

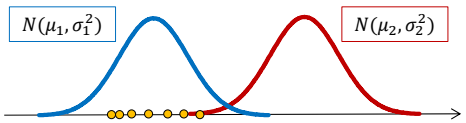
尤度: 与えられたデータがある確率分布から発生していると考えたときの尤もらしさの度合い

- データ y_i を発生させる確率分布: $p(y_i;\boldsymbol{\theta})$

  ($\boldsymbol{\theta} = [\theta_1, \dots, \theta_P]^T$)
- データ y_i の尤度: $l(\boldsymbol{\theta}|y_i) = p(y_i;\boldsymbol{\theta})$
- 独立同一分布から発生したデータ $\boldsymbol{y} = [y_1, \dots, y_N]^T$ の尤度:

$$l(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{N} p(y_i;\boldsymbol{\theta})$$

例: 次のデータは青と赤のどちらの正規分布から発生していると考えるのが妥当か？



$l_1(\mu_1, \sigma_1^2|\boldsymbol{y}) > l_2(\mu_2, \sigma_2^2|\boldsymbol{y})$

数学的準備 #2

データとパラメータの確率分布の表記方法

- $p(\boldsymbol{y})$: データベクトル $\boldsymbol{x} = [y_1, \dots, y_N]^T$ が生じる同時確率密度関数。$p(\boldsymbol{y}) = p(y_1, \dots, y_N)$
- $p(\boldsymbol{\theta})$: パラメータベクトル $\boldsymbol{\theta} = [\theta_1, \dots, \theta_P]^T$ の各要素を確率変数とみなしたときに $\boldsymbol{\theta}$ が生じる同時確率密度関数. $p(\boldsymbol{\theta}) = p(\theta_1, \dots, \theta_P)$

  例: 正規分布では $p(\boldsymbol{\theta}) = p(\mu, \sigma^2)$
- 確率の乗法定理: $p(\boldsymbol{y}, \boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$
- $p(\boldsymbol{y}) = \int p(\boldsymbol{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} = \int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}$
- $\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}$: パラメータの集合で積分をする記号

  例: $\theta_i \in \mathbb{R}\ (i = 1, \dots, P)$ なら,

  $\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\theta_1 \cdots d\theta_P$

データ&パラメータ&ベイズの定理 #1

データとパラメータを結びつけるベイズの定理

- **連続型確率変数のベイズの定理**

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

- $p(\boldsymbol{\theta}|\boldsymbol{y})$: **事後確率分布(事後分布)**
- $p(\boldsymbol{\theta})$: **事前確率分布(事前分布)**
- $p(\boldsymbol{y}|\boldsymbol{\theta})$: **尤度関数**
- $p(\boldsymbol{y})$: **周辺尤度**

データ&パラメータ&ベイズの定理 #2

尤度の観点からみたベイズの定理の意味

- **パラメータ θ を確率変数と見ている**
- **独立同一分布から発生したデータ $\boldsymbol{y} = [y_1, \ldots, y_N]^T$ を考えると, $p(\boldsymbol{y}|\boldsymbol{\theta})$ は尤度関数 $\prod_{i=1}^{N} p(y_i; \boldsymbol{\theta})$**
- **事後分布 $p(\boldsymbol{\theta}|\boldsymbol{y})$ はデータ観測後のパラメータの分布**

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\prod_{i=1}^{N} p(y_i; \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto \prod_{i=1}^{N} p(y_i; \boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

$\propto$: **比例するという記号**

- **ベイズの定理の構造**

事後分布 ⇐ データ(尤度関数) × 事前分布

演習問題

問題

- **ある企業では同じ型の製品を工場A,B,Cでそれぞれ生産し販売している。また, 不良品出荷率が少ないほど生産シェアを高く設定している。**
- **生産シェア　工場A:70%, 工場B:25%, 工場C:5%**
- **不良品出荷率　工場A:0.1%, 工場B:0.2%, 工場C:0.4%**

- **ある1つの販売済みの製品が不良品と判明した。その不良品が工場Aでつくられた製品である確率を求めなさい。**