特別講義シリーズ2023
因果推論の統計学的アプローチ

22FEB2024

統計的因果推論を理解するための基礎知識

東京大学大学院医学系研究科
公共健康医学専攻　生物統計学分野
松山　裕

1

3限の講義では

- 統計学、特に数理統計学について説明
  - 統計的推測に対するアプローチ：検定論と推定論
  - 推定論に焦点を絞る
    - 因果推論ではターゲットパラメータの推定（効果の大きさ）に関心がある
- 統計的因果推論で使用される手法の原理・原則の理解
  - 数理統計学の知識が必要
  - 推定方程式、ロバスト分散、二重ロバスト推定、…
- 複雑な因果解析手法をなぜ用いる必要があるのか?
  - 繰り返し測定がなされる治療（曝露）の効果に関心がある場面

2

2

そもそも“統計学”とは何か?

- 統計学で問題にするのは“バラツキ（Variation）”
  - バラツキとは
    - 同じようなものを測っているのに値がいろいろと違ってしまうこと、違っていること
- そのバラツキを
  - 誤差的な部分と意味（physical meaning）のある部分（対応可能部分）とに分解し、誤差を考慮に入れて意味を理解しようという視点・態度
- 2種類のバラツキ
  - バイアス　　　　：対応（説明）可能
  - ランダムエラー　：誤差と考える（みなす）
- バイアスとなりうるバラツキを確率論で評価可能な誤差に転化し、真実に迫ろうとする視点・態度
  - 方法：ランダムサンプリングとランダム化

3

3

統計的推測の3つの枠組み

- 現実の有限母集団からのランダムサンプリング
- 仮想的無限母集団からのランダムサンプリング
- 特定の標本に対するランダム化

例えば、リスク因子と疾患発症との間の関係を確率モデル（二項分布）により定式化し、その関係をモデルとして表現して、そのパラメータに関する推定や検定を行う

$$\log\frac{P}{1-P} = \beta_0+\beta_1(Age)+\beta_2(Gender)+\beta_3(Smoking)+\cdots$$

- 2番目の枠組み
  - 母集団が存在するかどうかにかかわらず、統計手法を用いるために考えられた枠組み
  - しかし、現実のほとんどのデータではランダムサンプリングを行っていない
    - 仮想的無限母集団と仮想的ランダムサンプリングという概念には無理がある
      - 同時代に生きる日本人全体、あるいは当該疾患を有する患者全体のような集団を仮想し、測定値をそれからのランダムな仮想標本の実現値と考える?
- データにモデルとして直接、確率分布を想定し、それを経験的に確かめる

4

4

Nonrandomly or nonrandomized sampled studyにおける

統計解析の3つのフェーズ：3S

- Summarization
  - データの要約・視覚化
  - 統計学的な議論（検定、信頼区間などの推測統計学）は必要ない
- Smoothing
  - 統計モデルの当てはめ
    - 統計モデル：データ生成過程（generating process）に対する仮定を数学的に表現したもの
    - 全ての統計手法はモデルであり、数学的な仮定がある
  - ターゲットパラメータの推定
    - 全てのモデルには興味のあるパラメータが含まれる
- Scientific inference
  - Smoothingの結果の説明・解釈
  - 純粋な統計学的なプロセスだけではない

Greenland. 1990, 1993.

5

5

統計モデル

- 2つのコンポーネントからなる
  - Y = f(X) + ε
    - f：アウトカムYとXの間の関係を規定する関数（structural component）
      - X: 説明変数（複数あってもよい）
    - ε：ランダム誤差（error component）
- 関数形 f とランダム誤差 ε に対する仮定
  - データ生成過程、研究目的などから統計学的・経験則的に決定
  - f の形
    - 統計解析で最も重要な特定

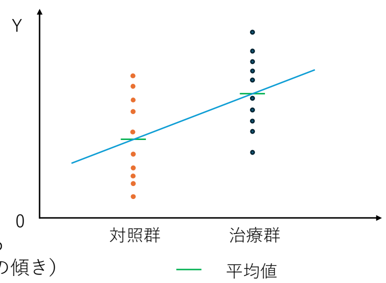
Robins and Greenland, 1986.

6

6

統計モデルの例1

- アウトカムYと治療変数Xの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
      - X: 治療群（1）、対照群（0）
    - ε：ランダム誤差
  - 関数 f の例
    - $Y = a_0 + a_1X + \varepsilon$
- パラメータ
  - 値が決まれば統計モデルが一意に定まる
    - ターゲットパラメータは a_1（回帰直線の傾き）
      - 平均値の差の t-検定
    - 局外パラメータ：特に興味がないパラメータ（a_0）

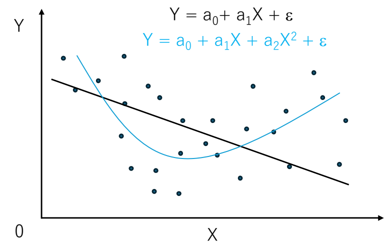


7

7

統計モデルの例2

- アウトカムYと説明変数Xの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
    - ε：ランダム誤差
  - 関数 f の例
    - $Y = a_0 + a_1X + \varepsilon$
    - $Y = a_0 + a_1X + a_2X^2 + \varepsilon$
- パラメータ推定の結果
  - 例えば、$E(Y) = 10 - 1.2X + 0.8X^2$



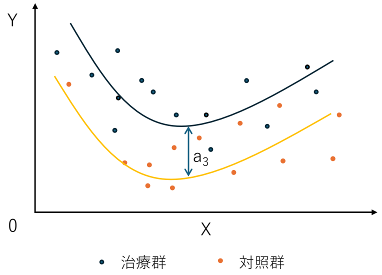
E(Y)：Yの期待値 Expectation

8

8

統計モデルの例3

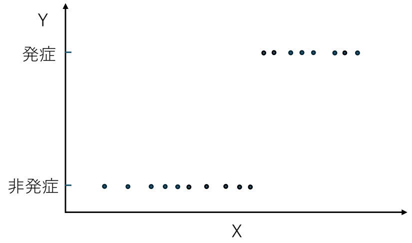
- アウトカムYと説明変数Xの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
    - ε：ランダム誤差
  - 関数 f の例
    - $Y = a_{0g} + a_{1g}X + a_{2g}X^2 + a_3\mathrm{group} + \varepsilon$
- ターゲットパラメータ
  - a_3：Xの影響を調整した群効果



9

統計モデルの例4

- アウトカムYと説明変数Xの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
    - ε：ランダム誤差
  - Y：二値データ
- 完全分離

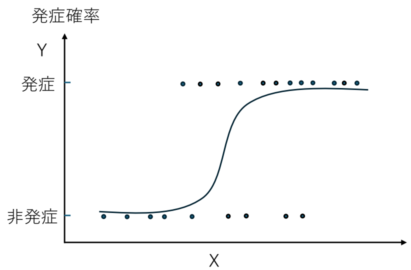


10

統計モデルの例4

- アウトカムYと説明変数Xの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
    - ε：ランダム誤差
  - Y：二値データ
- 関数 f の例
  - アウトカムの発症確率：E(Y)
    - $E(Y) = \frac{\exp(a_0 + a_1X)}{1+\exp(a_0 + a_1X)}$

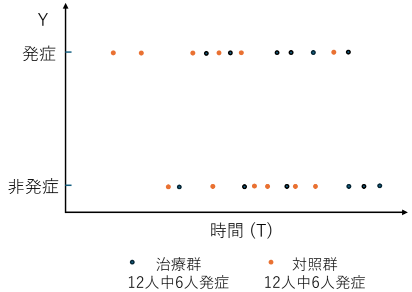
→ ロジスティック回帰：$log\frac{E(Y)}{1-E(Y)} = a_0 + a_1X$



11

統計モデルの例5

- アウトカムYと時間変数Tの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
    - ε：ランダム誤差
  - Y：二値データ
- 複合アウトカム
  - 単位時間当たりの発生数
    - 発症の速度（勢い）
  - アウトカムの発症ハザード（/時間）
    - hazard, incidence rate



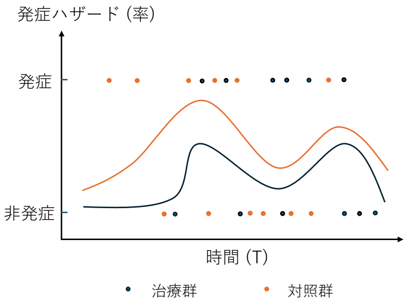
12

統計モデルの例5

- アウトカムYと時間変数Tの関係
  - Y = f(X) + ε
    - f：YとXの間の関係を規定する関数
    - ε：ランダム誤差
  - Y：二値データ
- 関数 f の例
  - 発症ハザード：h(t)
    - $h(t) = h_0(t)\exp(aX)$

$h_0(t)$：基準ハザード、X：治療群

→ 比例ハザードモデル：$\log h(t) = \log h_0(t) + aX$



13

13

パラメータ推定の問題に対する統計手法

- あらかじめ、ある統計量t(X)を用意しておき、実現値xが得られたら、それをXに代入し、得られるt(x)をパラメータ θ の真の値と判断
  - X：データ（確率変数）
  - x ：データの実現値（観察値）
- 推定量 Estimator
  - 推定に用いる方法（統計量）：t(X)
- 推定値 Estimate
  - 推定の結果得られる値：t(x)

ある集団に対する体重の真値?

・N個の確率変数：$X_1, \cdots, X_N$
・N人の対象者に体重測定
・実際の測定した結果： $x_1, \cdots, x_N$
・平均値で真値を推定

推定量：$\frac{\sum_{i=1}^{N} X_i}{N}$

推定値：$\frac{\sum_{i=1}^{N} x_i}{N}$

14

14

推定結果である推定値の良さの基準

- ある未知パラメータ θ に対する推定値 t(x)
- 二乗誤差 Squared error

$$\{t(x) - \theta\}^2$$

  - 推定値の真値からのズレの大きさ：小さいほうが良い
  - 実際には数値として得られない
  - 真値 θ はわからない（未知）

偏差（真値からのズレ）の大きさを絶対値 |t(x)-θ| で測ることもできるが、統計学的には良い指標ではない

15

15

現実的な対処方法

- 「方法としての良さ」を基準
- 推定
  - 方法：確率変数としての推定量：t(X)
  - 推定量に対する二乗誤差：これも確率変数（確率的に分布する）

$$\{t(X) - \theta\}^2$$

  - その分布がゼロの近くに集中している
- 平均二乗誤差 Mean squared error：MSE
  - 二乗誤差の分布の平均値（期待値）

$$E\{t(X) - \theta\}^2$$

16

16

平均二乗誤差を最小に

- 世の中、どんな面からみても良いものなどないのが普通
- どんな場合でも（未知パラメータの真値がパラメータ空間上のどこに存在しても）平均二乗誤差を最小にする推定量は存在しない

17

17

推定量の良さを計量的に定義

- <u>正確さ accuracy</u>
  - 不偏性 unbiasedness　$E[\hat{t}(X)] = \theta$　（期待値が真値に一致する）
  - 一致性 consistency　$\hat{t}(X) \underset{n\to\infty}{\to} \theta$　（nの増加とともに真値に収束する）
- <u>精密さ precision</u>
  - 分散最小　$Var[\hat{t}(X)]$ ができる限り小さい推定量のほうがよい
- 頑健さ robustness
  - 前提からのずれに敏感でない推定量
  - M-推定量、L-推定量、R-推定量
- 簡便さ
  - 計算が簡単な推定量のほうがよい

18

18

平均二乗誤差の分解

- 平均二乗誤差は、以下のように分解可能

$$E[t(X)-\theta]^2 = \underbrace{\{E[t(X)]-\theta\}^2}_{\text{推定量のバイアスの2乗}} + \underbrace{E[\{t(X)-E(t(X))\}^2]}_{\text{推定量の分散}}$$

バイアスと分散のトレードオフ

19

19

推定量の資格条件と定量的条件

- 資格条件：不偏性 Unbiasedness
  - 任意のθに対して、E[t(X)] = θ
  - 不偏推定量 Unbiased estimator
    - 推定量の分布のほぼ中心に未知パラメータの真値がある
  - 前頁の平均二乗誤差の第1項をゼロにする推定量
- 定量的条件：分散最小性 Minimum variance
  - Var[t(X)]を最小にする
    - 不偏性をもつ推定量のなかで、前頁の平均二乗誤差の第2項をなるべく小さくする
  - 不偏推定量のなかに、θがどんな値であっても分散が最小ものが存在するか？
    - あれば、一様最小分散不偏推定量（UMVU推定量）あるいは、最良不偏推定量（BU推定量）と呼ばれ、最も良い推定量

UMVU：uniformly minimum variance unbiased，BU：best unbiased

20

20

良い推定量の探し方

- 規準を満たすものを探す
  - 不偏推定量、最小分散、…
  - 規準を満たすものが無い場合もある
    - 特に、複雑な統計モデルにおいてはUMVU推定量は存在しない
- ある種の原理的な構成法を用意して、そこから推定量を導きだす
  - 導かれた推定量の性質・性能を調べ、その中で比較的良いものを採用

21

21

パラメータ推定に対する原理的な構成法

- 最小二乗法
- 最尤法
- モーメント法
- …

22

22

最小二乗法

- 最小二乗原理
  - 残差平方和を最小にするパラメータが良い
    - 残差 = 観測値 − モデルからの予測値
  - 回帰分析・分散分析といった線型モデルの場合にこの原理がうまく働く
    - Y = Xβ + ε,　β：回帰パラメータ、ε：ランダム誤差
- 関数形 f とランダム誤差に対する仮定
  - f：パラメータに関して線型（例えば、$Y = \beta_0+\beta_1X_1+\beta_2X_1^2+\cdots+\beta_pX_p$）で表現可能
  - ε：ガウス・マルコフ条件
    - E(ε) = 0　：不偏性（誤差の期待値はゼロ）
    - V(ε) = σ^2　：等分散性（誤差分散は一定）
    - 独立性　：誤差は互いに独立
    - 正規性の仮定は必要「なし」

23

23

最小二乗推定量： $\hat{\beta} = (X^T X)^{-1} X^T Y$

- 線型モデル
  - Y = Xβ + ε,　E(ε) = 0, V(ε) = σ^2
- 残差平方和を最小にする
  - $\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \;=\; \sum_{i=1}^{N}(Y_i - X_i\hat{\beta})^2 \;\rightarrow\; \min$
    - $\hat{Y}_i = E(Y_i) = X_i\hat{\beta}$
- 正規方程式
  - 細かい計算を省略し、行列・ベクトル表記すれば、

$$X^T X \hat{\beta} = X^T Y$$

24

24

最小二乗推定量 $\hat{\beta}=(X^TX)^{-1}X^TY$ の性質1

- $\hat{\beta}$ の期待値
  - $E(\hat{\beta})=\beta$
  - <u>不偏推定量（関数形 f が正しければ、バイアスはない）</u>
- $\hat{\beta}$ の分散
  - $V(\hat{\beta})=\sigma^2(X^TX)^{-1}$

$$\begin{aligned}E(\hat{\beta})&=E[(X^TX)^{-1}X^TY]\\&=(X^TX)^{-1}X^TE[Y] \quad \leftarrow E[Y]=X\beta\\&=(X^TX)^{-1}X^TX\beta\\&=\beta\end{aligned}$$

$$\begin{aligned}V(\hat{\beta})&=V[(X^TX)^{-1}X^TY]\\&=\underline{(X^TX)^{-1}X^TV[Y]\{(X^TX)^{-1}X^T\}^T}\end{aligned}$$

$V[Y]=\sigma^2(X^TX)^{-1}$ を代入すれば

ロバスト分散（サンドウィッチ分散）
（V[Y]の構造がどのような形であってもよい）

$$\begin{aligned}&=\sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}\\&=\sigma^2(X^TX)^{-1}\end{aligned}$$

25

25

最小二乗推定量 $\hat{\beta}=(X^TX)^{-1}X^TY$ の性質2

- 最小二乗推定量は不偏推定量
  - 推定量の分散の大きさが問題
  - $V(\hat{\beta})=\sigma^2(X^TX)^{-1}$
- 最良線型不偏推定量：Best Linear Unbiased Estimators(BLUE)
  - ガウス−マルコフの定理
    - <u>任意の線型不偏推定量の中で、最小分散を与える</u>

線型推定量：アウトカムYの線型結合で表現できる推定量

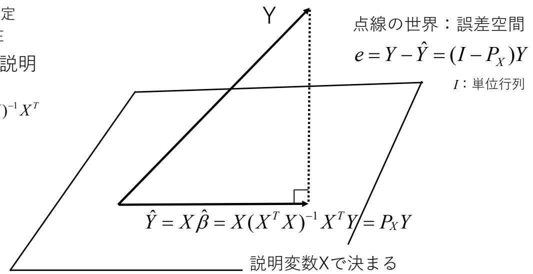
- 誤差の分布に<u>正規性 ε ~ N(0, σ2)</u> を仮定すると、
  - 最小二乗推定量は<u>UMVU推定量</u>（あらゆる不偏推定量の中で最小分散）

26

26

最小二乗推定量の幾何学的解釈
（YのXの上への回帰　Regression of Y on X）

- アウトカムベクトル：Y
  - N人の対象者に対して測定
    - N次元空間上に存在
- Yの平均的な挙動をXで説明
  - YのXへの正射影
  - 変換行列：$P_X=X(X^TX)^{-1}X^T$



27

27

ピタゴラスの定理：平方和の分解

- 前頁のベクトルYの始点を平均値($\bar{Y}$)とすれば、

$$\sum_{i=1}^{N}(Y_i-\bar{Y})^2 = \sum_{i=1}^{N}(\hat{Y}_i-\bar{Y})^2 + \sum_{i=1}^{N}(Y_i-\hat{Y}_i)^2$$

総平方和	群間平方和 モデル平方和	群内平方和 誤差（残差）平方和

- 誤差分布の正規性のもとでは、
  - F検定：コクランの定理を利用

$$F=\frac{\textbf{モデル平方和}/(p-1)}{\textbf{誤差平方和}/(N\text{-}p)}$$ ~ 自由度（p-1, N-p）のF分布に従う　p：パラメータの数

28

28

正規方程式 $X^T X\hat{\beta} = X^T Y$ を要素で書き表すと、

- 説明変数が1個の場合（パラメータ数は2個）
  - $Y_i = \beta_0 + X_i\beta_1 + \varepsilon_i$
- 方程式は2つ
  - $\sum_{i=1}^{N}[y_i - (\beta_0 + x_i\beta_1)] = 0$ ⟶ 残差の和がゼロになる推定量が良い
  - $\sum_{i=1}^{N}x_i[y_i - (\beta_0 + x_i\beta_1)] = 0$ ⟶ 説明変数と誤差の空間が直交する推定量が良い

一般に、2つのベクトルの要素を掛けて足したもの $\sum a_i b_i$ は内積 $\vec{a}^T\vec{b}$ と呼ばれる
内積がゼロならば、2つのベクトルは直交する

29

29

非線型なモデルの場合のパラメータ推定法

- 非線型モデル
  - ロジスティック回帰、ポアソン回帰、比例ハザード回帰など
  - 最小二乗法ではうまくパラメータ推定ができない
- 最尤推定法（ML: Maximum Likelihood）
  - 最尤原理：<u>尤度関数を最大化するパラメータが良い</u>
    - 誤差の正規性のもとでは、最小二乗推定量と最尤推定量は一致する

30

30

尤度（ゆうど）関数：Likelihood function L(θ; x)

- データ生成過程に対する確率モデル
  - 正規分布、二項分布、ポアソン分布、指数分布、…
- 例えば、N人の対象者のうちX人発症したとする
  - この状況を発症確率 θ の二項分布で表現する

$$L(\theta; x) = f(x; \theta) = {}_N C_x\, \theta^x (1-\theta)^{N-x}$$

関数g(a; b)：bが観測値でaがパラメータ

- 推定の問題
  - x: 測定値として既知、θ: 真の値が未知
  - θにはいろいろな値を考える必要がある（θを変数とみなす必要がある）
  - <u>確率関数 f(x; θ)をθの関数としてみる必要がある</u>

31

31

確率分布 f(x; θ)をθの関数とみると?

- 確率変数（データ） X
  - いろいろな値をとりえるが…
- 実現値 x
  - 関数 f(x; θ)が大きな値のところから実現しやすいはず
  - あるθの値に対して、f(x; θ)が大きな値ならば、そのようなθの値は真値の可能性が高い
- f(x; θ)をθの関数としてみた尤度関数
  - 各θの値が真値である可能性の目安
  - <u>尤度関数 L(θ; x) が最大となるθの値はデータを説明するのに尤もらしい</u>

32

32

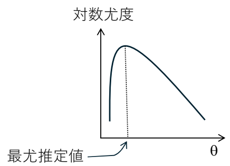
対数尤度の最大化

- 二項分布の例

$$L(\theta;x)=f(x;\theta)={}_N C_x\,\theta^x(1-\theta)^{N-x} \qquad \rightarrow \ \theta\text{に関して最大化}$$

- 対数尤度関数を最大化するほうが計算上便利
  - 対数関数の単調性から、Lが最大になるθでは、log(L)も最大になる

$$\log L(\theta;x) \ = \ \log {}_N C_x + x\log\theta + (N-x)\log(1-\theta) \qquad \rightarrow \ \theta\text{に関して最大化}$$



対数尤度関数は上に凸な二次関数
（指数型分布族の場合）

33

33

尤度方程式

- 対数尤度関数の最大化
  - 尤度方程式 Likelihood equation
  - 対数尤度関数を未知パラメータに関して微分して、ゼロとおいた式
  - 前頁の例では、

$$\frac{\partial \log L}{\partial\theta}=\frac{x}{\theta}-\frac{N-x}{1-\theta} \quad (=0)$$

$$\Longrightarrow \quad \hat{\theta}=\frac{x}{N}$$

34

34

推定方程式

- 最尤推定量を求めるためには、
  - ランダム誤差（error component）に対する確率モデルを仮定
  - その分布を未知パラメータの関数としてみた対数尤度関数を最大化
    - 尤度方程式：対数尤度関数を未知パラメータに関して1回微分してゼロとおいた式
- スコア関数

$$S(\theta)=\frac{\partial \log L(\theta;X)}{\partial\theta}$$

  - 推定方程式：スコア関数=0
    - 尤度方程式と同じだが、尤度が存在するかどうかは別としてこの方程式を解けばパラメータの解が得られるという意味
    - 誤差モデルが正規分布の場合は、正規方程式に一致

35

35

スコア関数の性質 $S(\theta)=\frac{\partial \log L(\theta;X)}{\partial\theta}$

- スコア関数は標本Xの関数であるので、それ自体が確率変数
  - 期待値や分散などのモーメントをもつ
- スコア関数の期待値
  - E[S(θ)] = 0 ⟶ 期待値は常にゼロ
- スコア関数の分散
  - V[S(θ)] = E[S(θ)²] = I(θ) ⟶ 分散はフィッシャー情報量
    - I(θ)：フィッシャー情報量

$$I(\theta)=E\left[-\frac{\partial^2}{\partial\theta^2}\log L(\theta;X)\right]$$

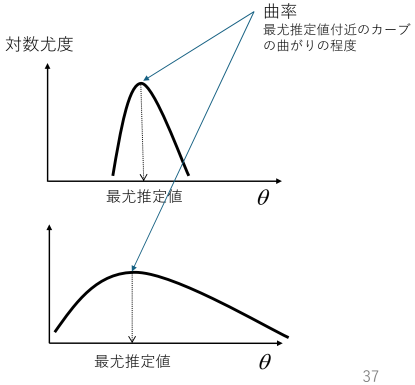
36

36

フィッシャー情報量：I(θ)

$$I(\theta)=E\left[-\frac{\partial^2}{\partial\theta^2}\log L(\theta;X)\right]$$

- データがもつ情報量
- 情報量が大
  - 推定精度が高い（分散が小さい）
  - $1/I(\theta)$ → 小さくなる
- 情報量が小
  - 推定精度が低い（分散が大きい）
  - $1/I(\theta)$ → 大きくなる



37

37

ちなみに、正規分布の場合

$$X_1, X_2, \cdots, X_n \sim N(\mu,\sigma^2)$$

$$L=f(x;\mu,\sigma^2)$$
$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{(x_i-\mu)^2}{2\sigma^2}] = (2\pi\sigma^2)^{-n/2}\exp[-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}]$$

- フィッシャー情報量 I(θ)
  - 平均値 μ について考えると、標本平均 $\bar{X}$ は不偏推定量
  - $\bar{X}$ の分散 $V(\bar{X})$ は?
    - 教科書には、 $V(\bar{X})=\frac{\sigma^2}{n}$

$$I(\theta) = E[-\frac{\partial^2 \log f(x\mid\mu)}{\partial\mu^2}]=\frac{n}{\sigma^2}$$

フィッシャー情報量の逆数

- 線型モデル Y = Xβ + ε, ε~N(0, σ^2) の場合
  - β に関するフィッシャー情報量 $\frac{(X^TX)}{\sigma^2}$
  - 逆数をとると、$\sigma^2(X^TX)^{-1}$

X = (1,1,...,1)ᵀの場合

$$X^TX=(1\ \ 1\ \ \cdots\ \ 1)\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}=n$$

38

38

スコア関数 $S(\theta)=\frac{\partial\log L(\theta;X)}{\partial\theta}$ の近似

x=aの周りでのテイラー展開

$$f(x)=f(a)+(x-a)f'(a)+(x-a)^2\frac{f''(a)}{2!}+\cdots$$

- スコア関数を真値 θ の周りでテイラー展開
  - 1次近似

$\hat{\theta}_{ML}$：最尤推定量

$$S(\hat{\theta}_{ML})\approx S(\theta)+\frac{\partial}{\partial\theta}S(\theta)(\hat{\theta}_{ML}-\theta)\quad(=0)$$

$$\rightarrow \boxed{S(\theta)\approx I(\theta)(\hat{\theta}_{ML}-\theta)}$$

大数の法則から

$$-\frac{\partial}{\partial\theta}S(\theta)=-\frac{\partial^2\log L(\theta;X)}{\partial\theta^2}\approx I(\theta)$$

- スコア関数の性質
  - 期待値はゼロ
  - 分散はフィッシャー情報量 $I(\theta)$

39

39

最尤推定量の3つの性質

- θ の最尤推定量は、漸近的に正規分布 N(θ , I^{-1}(θ))に従う
- 一致性 Consistency
- 漸近有効性 Asymptotic efficiency
- 漸近正規性 Asymptotic normality

40

40

一致性 Consistency

- ある推定量 $\hat{\theta}(X)$ が真値に確率収束する

$$\hat{\theta}(X) \xrightarrow{p} \theta \quad (n \to \infty)$$

- 標本数が多いとき、<u>最尤推定量は”ほぼ”不偏推定量</u>
  - 不偏性：期待値が真値に一致する
  - スコア関数の期待値 E[S(θ)] = 0

$$S(\theta) \approx I(\theta)(\hat{\theta}_{ML} - \theta)$$

41

41

漸近有効性 Asymptotic efficiency

- 最尤推定量の分散
  - 真値の周りにどれくらいのバラツキを示すか?
- スコア関数の分散: I(θ)

$$S(\theta) \approx I(\theta)(\hat{\theta}_{ML} - \theta)$$

両辺の分散を考えると

$$\to I(\theta) = I(\theta)^2 \,\mathrm{Var}(\hat{\theta}_{ML})$$

$$\to \mathrm{Var}(\hat{\theta}_{ML}) = 1 / I(\theta)$$

最尤推定量の分散はフィッシャー情報量の逆数

42

42

有効推定量とは?

- クラメル・ラオの不等式
  - 任意の不偏推定量 $\hat{\theta}(X)$ の分散の下限はフィッシャー情報量の逆数

$$\mathrm{Var}(\hat{\theta}(X)) \geq \frac{1}{I(\theta)}$$

- 最尤推定量の漸近分散はクラメル・ラオの下限を達成
- そのような推定量: 有効推定量 Efficient estimator
- Nが大きいとき、<u>最尤推定量は有効推定量</u>
  - すなわち、<u>ほぼ UMVU推定量</u>

43

43

漸近正規性 Asymptotic normality

- f(x; θ)からの大きさnの独立なデータ $X_1, X_2, \cdots, X_n$
- このとき、スコア関数は互いに独立に同一分布に従う確率変数の和

$$S(\theta; X) = \sum_{i=1}^{n} S(\theta; X_i)$$

  - $S(\theta; X_i) = \frac{\partial}{\partial \theta} \log f(X_i; \theta)$

- 中心極限定理より、独立な確率変数の和は漸近的に正規分布に従う
- 漸近的に $(\hat{\theta}_{ML} - \theta)$ はスコア関数の線型関数となる

$$(\hat{\theta}_{ML} - \theta) \approx I(\theta)^{-1} S(\theta)$$

44

44

最尤推定量の3つの性質（再掲）

- θ の最尤推定量は、漸近的に正規分布 N(θ, $\mathrm{I}^{-1}(\theta)$)に従う
- 一致性 Consistency
- 漸近有効性 Asymptotic efficiency
- 漸近正規性 Asymptotic normality

45

45

ロジスティック回帰では、

- ロジスティック回帰
  - アウトカムが二値データの場合に頻用される統計モデル
  - データ生成過程：二項分布を想定するのが自然
- 尤度関数
  - $L(\theta;x) = {}_N C_x \, \theta^x (1-\theta)^{N-x}$
- 未知パラメータ θ に対する統計モデル
  - 説明変数が1個の場合（モデルのパラメータは2個）

$$\theta = \frac{\exp(\beta_0 + \beta_1 x)}{1+\exp(\beta_0 + \beta_1 x)}$$

46

46

推定方程式

- 尤度関数 $L(\theta) = \prod_i \theta_i^{y_i} (1-\theta_i)^{1-y_i}$
  - y_i：i番目の対象者が発症すれば1、非発症なら0
- 対数尤度関数

$$\theta_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)}$$

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \{ y_i \log \theta_i + (1-y_i) \log(1-\theta_i) \}$$
$$= \sum_{i=1}^n \{ y_i \beta_0 + x_i y_i \beta_1 - \log(1+e^{\beta_0+\beta_1 x_i}) \}$$

- 未知パラメータβ_0とβ_1について偏微分

$$\sum_{i=1}^n \left(y_i - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right) = 0$$
$$\sum_{i=1}^n x_i \left(y_i - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right) = 0$$

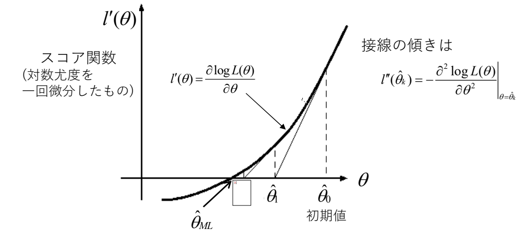
・残差の和がゼロ
・残差空間と説明変数空間の直交性

47

47

推定方程式が非線形

- 方程式の明示的な解がない
- 反復計算アルゴリズムで解く
  - ニュートン・ラフソン法などの数値計算



48

48

推定問題での統計学的議論　推定値と信頼区間幅

適切なターゲットパラメータを設定し、その推定量の
- バイアス
  - 仮定したモデルが正しいという前提で、不偏推定量（あるいは一致推定量）を与える方法
  - 資格条件
- 分散
  - バイアスのない推定ができているという前提で、できる限り小さな分散を与える方法
  - 定量的条件
    - UMVU推定量、最尤推定量
    - できる限り検出力が高い方法を使うべき

49

49

いくつかの原因によるバイアス

- 比較におけるバイアス
  - 交絡バイアス
  - 重要な交絡因子を計画段階から同定し、収集する
- 測定におけるバイアス
  - 測定誤差バイアス
    - 曝露、アウトカムの誤分類：Differential/Non-differential misclassification
- 追跡におけるバイアス
  - Attritionバイアス（情報のある脱落）
- 解析におけるバイアス
  - モデル（関数形）の誤特定

Greenland. 1997, Greenland S, VanderWeele. 2015.

50

50

因果解析のための手法にバイアスがないためには

- いくつかの仮定がある
- なかでも、
  - 未測定の交絡因子がない
    - 条件付き独立（交換可能性）の仮定
  - 解析に使用したモデルが正しい

51

51

因果効果の推定

- Point treatment (exposure)
  - 治療（曝露）もアウトカムも1回のみ測定
    - 標準的な統計解析手法がある
      - 平均値の差、リスク差、リスク比、オッズ比、ハザード比を推定するための手法
    - 因果解析手法も使用できる
  - 治療は1回だけ測定（変化しない）が、アウトカムを複数回測定
    - 標準的な統計解析手法がある
      - 反復測定解析：混合効果モデル、一般化推定方程式（GEE）
    - 因果解析手法も使用できる
- Time-varying treatment (exposure)
  - 治療を複数回測定（治療内容が変化する）
    - 上記の標準的な手法では効果推定値にバイアスがある
    - 因果解析手法を使用する必要がある

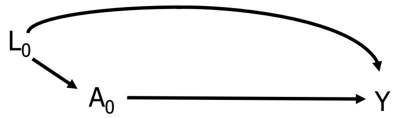
52

52

Point treatmentの場合

- A：治療変数、Y：アウトカム変数、L：交絡変数
  - 交絡因子Lによるバイアスを補正した治療効果の推定

A,L,Yがすべて二値データの非常に単純な状況
・標準的な解析と因果解析の推定結果は同じ



  - 回帰・分散分析、ロジスティック回帰、Cox回帰などで十分では?
  - 上記のモデル（関数 f）が近似的に正しいもとでは、最も小さな分散を与える
- 傾向スコアを利用した因果解析
  - 利点：傾向スコアを正しく推定できれば、モデル(関数 f)を仮定する必要がない
  - 欠点：検出力が低い（分散の値が大きい） ⟶ 二重ロバスト推定量（次回のセミナー）

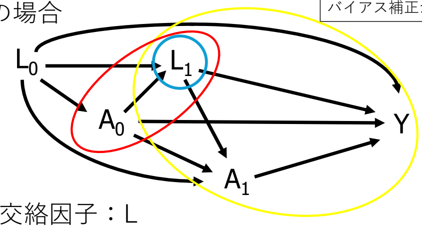
53

53

Time-varying treatmentの場合

- 治療が2回の場合

通常の層別解析・回帰モデル・傾向スコア法では
バイアス補正ができない



- 時間依存性交絡因子：L
  - 治療によって変化する中間結果でもあり、次の治療を決める因子
  - 交絡因子なので、調整しないと治療効果の推定にバイアス
  - 中間結果なので、調整すると治療効果の推定にバイアス

54

54

反事実アウトカムを利用した因果解析手法

- Robinsのg-methods
  - G-computation
  - Marginal Structural Models (MSMs)
  - Structural Nested Models（SNMs）

A,L,Yがすべて二値データの非常に単純な状況（関数形 f が飽和モデルの場合）
・3つの方法による推定結果は同じ
・推定値の分散が異なる

55

55

Generalized treatment regimes

- 治療が1回であれば
  - 治療を受けるか/受けないか
    - ある人が治療を受けた場合/受けなかった場合の結果
- 治療が複数回になると治療レジメンが複数存在
  - 最初から最後まで治療を受け続ける/受け続けない
  - ある検査値が閾値を超えたら治療を受ける/受けない
  - 副作用が発現すれば治療中止、回復すれば治療開始
  - …
  - K回治療を受ける場合、共変量に依存せずに決めたとしても 2^K 通りある

56

56

標準的な手法との関係

- G-computation
  - モデル（関数 f）を前提としない”ノンパラメトリック”な 方法
  - SMRなどの標準化 (Standardization) 法の拡張
- Marginal Structural Models (MSMs)
  - 傾向スコアを利用した方法
  - 標準的な回帰モデル法の拡張
- Structural Nested Models（SNMs）
  - 傾向スコアと構造モデルを利用した方法
  - 操作変数法の拡張

57

57

標準的な手法の拡張といっても…

- 最小二乗原理、尤度原理は利用できない
  - 尤度関数が構成が難しい
  - 複数の治療変数、時間依存性交絡因子を取り扱うためのモデルが必要
- 方法の理論的正当化をするためには<u>セミパラメトリック理論</u>が必要
  - 推定方程式と二重ロバスト推定

$$
\begin{aligned}
U(\mu_1) &= \sum_i \left[\frac{x_i(Y_i - \mu_1)}{PS_i} - \frac{x_i - PS_i}{PS_i}[RM_i(set\ x_i = 1) - \mu_1] \right] \\
&= \sum_i \left[\frac{x_i[Y_i - RM_i(set\ x_i = 1)]}{PS_i} \right] + \sum_i [RM_i(set\ x_i = 1) - \mu_1] \qquad (= 0)
\end{aligned}
$$

PS：傾向スコア、RM:回帰分析による予測値

58

58

“統計学”の理論・理屈を理解することは重要

- 新しい方法論の開発
  - 型どおりでない場面への対処
- 解析結果（コンピュータの出力結果）の正しい理解
- 他人への結果の説明
- 他人の主張・論文の理解と批判

59

59

参考文献

- 竹村彰通．現代数理統計学．学術図書出版社．2020.
- Greenland S Randomization, statistics, and causal inference. Epidemiology 1990; 1: 421-9.
- Greenland S. Summarization, smoothing, and inference in epidemiologic analysis. Scand J Soc Med 1993; 21(4): 227-32.
- Greenland S. Concepts of validity in epidemiological research. In Oxford Textbook of Public Health. PP 621-39. 1997.
- Greenland S, VanderWeele TJ. Validity and bias in epidemiological research. In Oxford Textbook of Global Public Health (6th edn). PP 569-90. 2015.
- Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. Am J Epidemiol 1986; 123(3): 392-402.
- Tsiatis AA. Semiparametric Theory and Missing Data. Springer: New York, 2006.

60

60