

数理統計 補助資料 ～統計モデルと尤度原理～

2024年度2学期: 月曜1限, 水曜3限
担当教員: 石垣 司

1

統計モデルとは？

確率モデル

- 確率現象を数学(確率論)で記述したモデル
標本空間, 確率変数を含む確率分布などを指す場合が多い

統計モデル

- 確率的構造を伴う対象や現象の数理モデル
現象を説明, 予測, 制御するための仮説の表現
実証, 予測, 制御など目的に応じたモデルが必要
線形回帰モデルも統計モデルの一種

統計的モデリング

- より良い統計モデルを作る・探求する行為

2

統計科学とモデル

赤池弘次『時系列解析の心構え』時系列解析の実際(赤池、北川編)朝倉書店 1995

統計的方法の本質は、データを用いて必要な情報を創り出すことにある。(中略) 統計科学は、このような情報の創出の過程に関する人間の知的活動の科学と定義される

モデルは仮説の表現であり、仮説の提案こそが基本的な知的活動である。これによって、データに意味が発生し、情報が創り出されるのである



Google を検索または URL を入力

2017年11月5日のGoogleロゴ

<https://www.google.com/doodles/hirotugu-akaikes-90th-birthday?hl=ja>

3

統計モデルに関する授業での問題意識

世の中のデータ分析は線形回帰モデルだけで十分か？

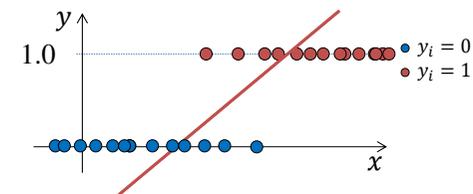
- 例: 目的変数 y が2値の離散変数 $\{0,1\}$ の回帰分析

【説明変数】

毎日の喫煙量
広告視聴時間
毎日の学習時間
気温の累計

【目的変数】

肺がんの発病 or 非発病
商品を買う or 買わない
志望校への合格 or 不合格
植物が発芽する or しない



線形回帰モデル(直線)ではデータの傾向をうまくとらえることができない

4

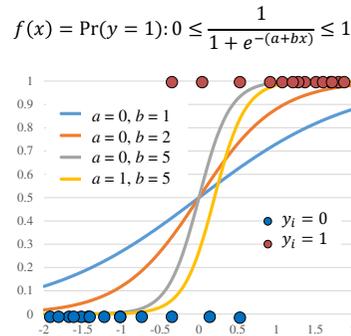
モデルの形状の改良が必要な例

例：ロジスティック回帰モデル

– 例：消費者による商品の購買/非購買の選択

– 関数形： $f(x) = \frac{1}{1+e^{-(a+bx)}}$

「推定」とはデータをうまく表現できるS字カーブの形状を決めること



【抽出できる情報の例】

商品Aが250円の時の購買確率70% (30%非購買)
商品Aが300円の時の購買確率20% (80%非購買)

線形回帰分析の関数形(回帰直線)ではデータにうまくフィットしないS字型の関数形が必要。

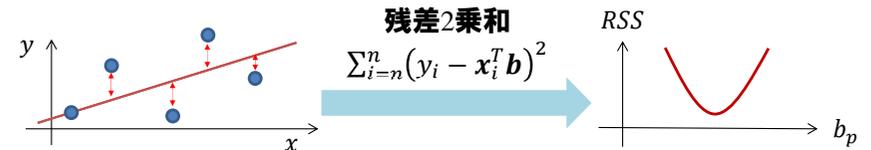
最小2乗法によるパラメータの推定

最小2乗法の原理

– データとモデルの誤差($y_i - \hat{y}_i$)の残差2乗和を最小化

– 考え方：モデルとデータの全体的なズレは小さい方が良い

線形回帰モデルの残差2乗和は b_p の2次関数のため、(正則ならば)最小点が存在し、かつ、ただ一つに決まる



– 一方、ズレの大きさの定義に確率的概念が含まれていない

◎ 確率を考える必要がないのでより簡単な表現かつ直感的

× パラメータ推定に関して確率を介したより一般的な議論ができない

尤度に基づくパラメータ推定

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： $p(y_i; \theta)$

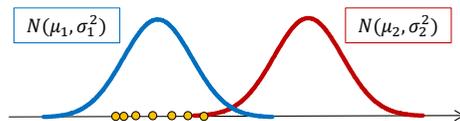
(パラメータ $\theta = [\theta_1, \dots, \theta_p]^T$)

– データ y_i の尤度： $l(\theta|y_i) = p(y_i; \theta)$

– 独立同一分布から発生したデータ $y = [y_1, \dots, y_N]^T$ の尤度：

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

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？



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

確率分布や尤度の表記法の確認

$p(y)$ ：確率変数 y の確率分布：

– この授業の範囲内では、具体的な確率関数が確率密度関数

$p(y, z)$ ：確率変数 y と z の同時分布(結合分布)

$p(y|z)$ ：確率変数 z の値が定まったときの y の条件付き分布

– $p(y, z)$ は2変量の分布で、 $p(y|z)$ は1変量の分布

$p(y; \theta)$ ：パラメータ θ を持つ確率変数 y の確率分布

– ここでの「;」は、パラメータは確率変数ではない点を強調している

$l(\theta|y)$ ：確率変数 y の値が定まったときのパラメータ θ の関数

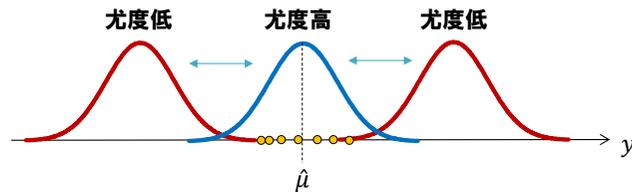
– 具体的な数式は確率関数・確率密度関数 $p(y; \theta)$ と同じ場合がほとんどである。ただし、引数が異なることに注意

最尤推定

尤度を最大にするパラメータを推定値とする推定方法

最尤推定の原理：尤度関数の最大化

- 尤度関数：尤度 $l(\theta|y)$ はパラメータ θ の関数
- 考え方：最も尤もらしいモデルからデータは発生する
- 例：分散 σ^2 の値を固定して正規分布 $N(\mu, \sigma^2)$ の μ を動かすと尤度 $l(\mu|y)$ の値は変化する。その中で尤度が最大となる μ を最尤推定値 $\hat{\mu}$ とする



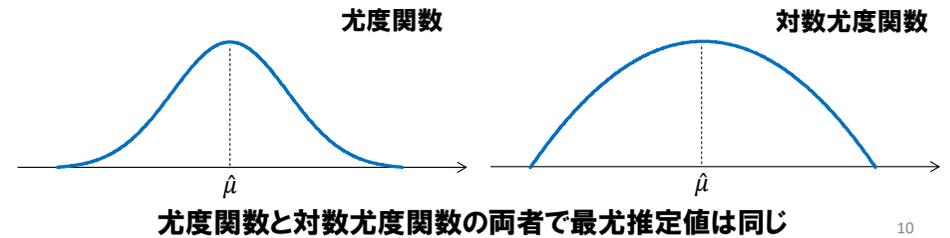
9

対数尤度関数

多くの場合、対数尤度関数を最適化に用いる

$$L(\theta|y) = \log\{l(\theta|y)\} = \sum_{i=1}^N \log\{p(y_i; \theta)\}$$

- 対数尤度関数は凸関数となることが多く使いやすい
- $L(\theta|y)$ が単峰性の関数のとき $\frac{\partial L(\theta|y)}{\partial \theta} = 0$ となるを探す
- 例：正規分布の平均 μ の尤度関数と対数尤度関数



10

具体例：正規分布の最尤推定

正規分布の尤度関数

- 確率密度関数： $f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\}$
- 尤度関数： $l(\mu, \sigma^2|y) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}$

正規分布のパラメータの最尤推定量

- 独立同一分布から発生したデータ $y = \{y_1, \dots, y_N\}$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N y_i$$

check!

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\mu})^2$$

#メモ：尤度を原理とした最適化を行っても正規分布の平均の推定量は最小分散線形不偏推定量と一致する。しかし、分散の推定量は一致しない。最尤推定量は不偏性を持つとは限らないので注意

11

具体例：回帰係数の最尤推定

線形回帰モデルの尤度関数

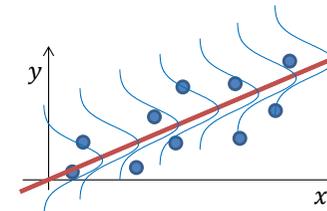
- 線形重回帰モデル： $y = Xb + e, e \sim N(0, \sigma^2 I)$
 $N(0, \sigma^2 I)$ は N 次元多変量正規分布
 $x_i = [1 \quad x_{i1} \quad \dots \quad x_{ip}]^T$ (行列 X の第 i 行の転置ベクトル)
- 尤度関数： $l(b|D) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i^T b)^2\right\}$

回帰係数の最尤推定量

$$\hat{b} = (X^T X)^{-1} X^T y$$

check!

最小2乗推定量と一致。原理が異なってもある仮定の下では同じ結果が得られるのは安心できる



$$y = \hat{b}_1 x + \hat{b}_0$$

- : 最尤推定された回帰直線
- : 各 x での尤度関数

12

演習問題

1. 正規分布の平均と分散の最尤推定量をそれぞれ導出
しなさい
2. 前のページの線形回帰モデルの回帰係数の最尤推定
量を導出しなさい
 - ベクトルとベクトルの積の関係: $a^T x = x^T a$
 - $X^T X = \sum_{i=1}^N x_i x_i^T$
 - ベクトルの微分の公式1: $f(x) = a^T x$ のとき $\frac{df(x)}{dx} = a$
 - ベクトルの微分の公式2: A が対称行列のとき
 $f(x) = x^T A x$ のとき $\frac{df(x)}{dx} = 2Ax$