数理統計 補助資料
～統計モデルと尤度原理～

2023年度2学期:　月曜1限, 水曜3限

担当教員:　石垣　司

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
  - 仮説の良さをデータから実証
  - 実データ分析のための回帰モデルは統計モデル

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

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

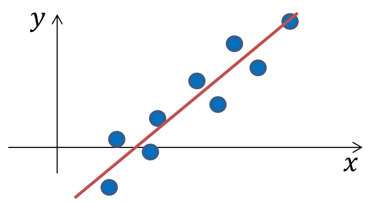
モデルは仮説の表現であり, 仮説の提案こそが基本的な知的活動である。これによって, データに意味が発生し, 情報が創り出されるのである

2017年11月5日のGoogleロゴ
https://www.google.com/doodles/hirotugu-akaikes-90th-birthday?hl=ja

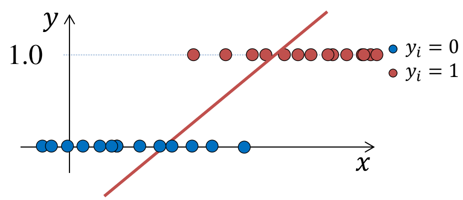
統計モデルに関する授業での問題意識

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値変数の場合

【説明変数】	【目的変数】
毎日の喫煙量	肺がんの発病 or 非発病
広告視聴時間	商品を買う or 買わない
毎日の学習時間	志望校への合格 or 不合格
気温の累計	植物が発芽する or しない



線形回帰モデルでOK!



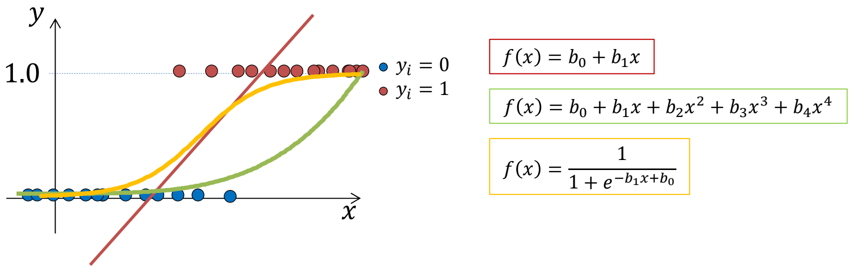
線形回帰モデルでOK???

統計モデルを利用するために

データに適した統計モデルを利用する

– モデル：関数 $f(x)$ と確率的構造を組み合わせる

例：線形単回帰モデル $f(x) = a + bx, y_i = f(x) + \varepsilon_i, \varepsilon_i \sim N(\mu, \sigma^2)$



どんな形の $f(x)$ と ε の確率分布を考えてもよい。ただし，合理的にパラメータを決める必要がある

最小2乗法によるパラメータの推定

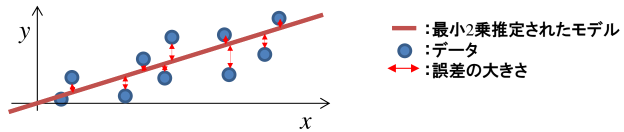
最小2乗法の原理

– モデルとデータの誤差の2乗和（RSS）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RSSはパラメータの2次関数

そのため，（正則ならば）最小点が存在し，かつ，ただ一つに決まる



– しかし，一般の $f(x)$ のRSSが2次関数になるとは限らない

もしも，1次や3次の項があると最小点はマイナス無限大に発散

#メモ：関数形を決めるだけなら、誤差項に正規分布の仮定は不必要。計量経済学の講義で学習する回帰係数の統計的検定や推定の不偏性・一致性などの推定量の性質は、誤差項に確率分布を仮定したときの論理展開

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：$p(y_i; \boldsymbol{\theta})$

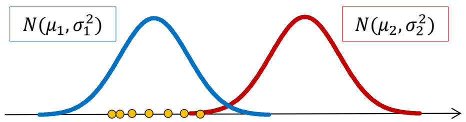
（パラメータ $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]^T$）

– 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 $\boldsymbol{y} = [y_1, \ldots, y_N]^T$ 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？



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

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，具体的な確率関数か確率密度関数

$p(y, z)$：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

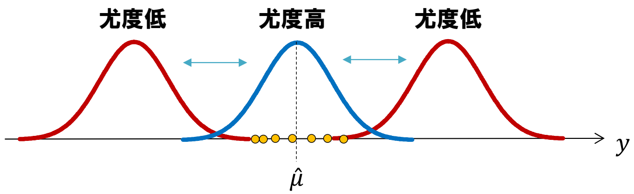
最尤推定

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 $l(\theta|\boldsymbol{y})$ 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 $l(\mu|\boldsymbol{y})$ 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。



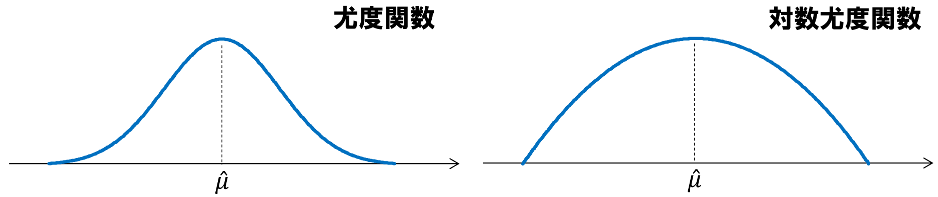
対数尤度関数

多くの場合，対数尤度関数を最適化に用いる

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

- 対数尤度関数は凸関数となることが多く使いやすい

  例：正規分布の平均パラメータ μ の尤度関数と対数尤度関数



尤度関数と対数尤度関数の両者で最尤推定値は同じ

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：$l(\mu, \sigma^2|\boldsymbol{y}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i-\mu)^2\right\}$

正規分布のパラメータの最尤推定量

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

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。

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：$\boldsymbol{y} = X\boldsymbol{b} + \boldsymbol{e}, \boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$

  $N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ は N 次元多変量正規分布

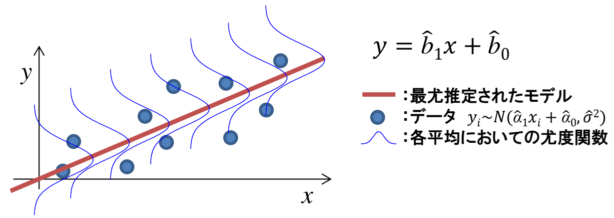
  $\boldsymbol{x}_i = [1 \quad x_{i1} \quad \cdots \quad x_{iP}]^T$（行列 X の第 i 行の転置ベクトル）
- 尤度関数：$l(\boldsymbol{b}|D) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{b}\right)^2\right\}$

回帰係数の最尤推定量

check!

$$\widehat{\boldsymbol{b}} = (X^TX)^{-1}X^T\boldsymbol{y}$$

#メモ：最小2乗推定量と一致



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. 正規分布 $N(y;\mu,\sigma^2)$ と指数分布 $f(y;\lambda)$ の確率密度関数はそれぞれ次式で与えられる。

$$N(y;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}$$

$$f(x;\lambda)=\lambda\exp(-\lambda y)$$

独立同一分布から発生したデータ $\boldsymbol{y}=[1,2,3]^T$と

$N\left(\boldsymbol{y};\mu=1,\sigma^2=\frac{1}{2\pi}\right)$と $f(\boldsymbol{y};\lambda=1)$ の尤度を

それぞれ求めなさい