

<https://goo.gl/tYmyZL>

KYOTO UNIVERSITY

統計的モデリング基礎④ ～最尤推定～

鹿島久嗣
(情報学科 計算機科学コース)

DEPARTMENT OF INTELLIGENCE SCIENCE
AND TECHNOLOGY

中間テスト：

6/7(水) 4限 総合研究8号館NSホール

- 講義内で行います
- 範囲は第1～7回
- 持ち込み等なし

(いろいろな話題についての) 参考書

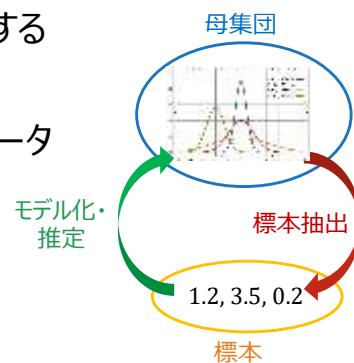


現代統計学 (2017)

発刊年月: 2017.03
ISBN: 978-4-535-78818-3
A5判; 256ページ

統計モデリングの考え方： 部分から全体について知る

- 母集団：
確率分布で表される、本当に興味のある集合
– 分布のクラスやパラメータで指定されるとする
- 標本：
確率分布に従って抽出された具体的なデータ
- 目的：
標本から母集団について推測する
(標本抽出の逆)
– パラメータを推定する (どうやって?)



パラメータの推定問題：

サイコロの各目の出る確率を実際の出目から推定する

- 母集団は離散分布に従うとする

– $P(X = k) = f(k)$ (ただし $\sum_{k \in X} f(k) = 1$)

– たとえば (厳密な) サイコロであれば $P(X = k) = \frac{1}{6} \approx 0.17$

- 標本抽出：

– 20回 (独立に) 振ったところ、

6 3 5 1 3 1 4 1 2 2 6 1 2 2 5 4 4 4 6 5 が出た

出目	1	2	3	4	5	6
回数	4	4	2	4	3	3

- 母集団のパラメータ (それぞれの目の出る確率) は？

5

KYOTO UNIVERSITY

サイコロの推定問題へのひとつの解：

割合で推定する

- ひとつのアイデア：

20回中で1が4回出たのだから $P(X = 1) \approx \frac{4}{20} = 0.2$ と推定する

出目	1	2	3	4	5	6
回数	4	4	2	4	3	3
確率の推定値	0.2	0.2	0.1	0.2	0.15	0.15

- 正解が約0.17なので悪くない...
- この推定値はどのような原理に基づいているのか？

6

KYOTO UNIVERSITY

最尤推定： 確率分布の代表的な推定手法のひとつ

- 標本からの母集団確率分布の推定
- 代表的な推定手法
 - 最尤推定
 - モーメント推定
 - ベイズ推定
 - ...

7

KYOTO UNIVERSITY

最尤推定とは： 標本をもっともよく再現するパラメータを推定値とする

- n 個のデータ： x_1, x_2, \dots, x_n が生成される確率（尤度）：

$$L = P(X = x_1)P(X = x_2) \cdots P(X = x_n) = \prod_{i=1}^n P(X = x_i)$$

- サイコロの例：

– 目 k が出る確率を p_k , 目 k が出た回数を n_k とする

– 尤度 $L(p_1, p_2, \dots, p_n) = p_1^{n_1} p_2^{n_2} \cdots p_6^{n_6} = \prod_{k=1}^6 p_k^{n_k}$

– これを最大化する p_1, p_2, \dots, p_n を求める（最大化問題を解く）と

$$\hat{p}_k = \frac{n_k}{n_1 + n_2 + \cdots + n_6}$$

8

KYOTO UNIVERSITY

サイコロの最尤推定： ラグランジュの未定乗数法

- 尤度の代わりに対数尤度を最大化すると扱いやすい（解は変わらない）：

$$\log L(p_1, p_2, \dots, p_n) = \sum_{k=1}^6 n_k \log p_k$$

- 確率分布の制約: $\sum_{k=1}^6 p_k = 1, p_k > 0$
- ラグランジュ未定乗数法：

$$G(\{p_k\}_{k=1}^6, \lambda) = \sum_{k=1}^6 n_k \log p_k + \lambda \left(1 - \sum_{k=1}^6 p_k \right)$$

問題： どちらのサイコロが使われた？

- 2つの（いびつな）サイコロA, Bがある

–サイコロAを20回振ったところ：

出目	1	2	3	4	5	6
回数	5	1	4	2	4	4

–サイコロBを16回振ったところ：

出目	1	2	3	4	5	6
回数	2	8	2	2	1	1

- 2つのサイコロのいずれかを選んで（Cとする）5回振ったところ：

出目	1	2	3	4	5	6
回数	1	1	0	2	0	1

- 使われたサイコロはA, Bのいずれだろうか？（C=A or C=B?）

ベイズ決定： 事後確率によって決定する

- A, B どちらのサイコロを選んだかを確率変数 X で表す
 - 事前確率：でたらめに選ぶと $P(X = A) = P(X = B) = 1/2$
 - 何も情報がなければこれ以上はわからない
- 事後分布：C (A, Bのいずれか) 振って出たデータ D を見たあとの X の確率分布 $P(X|D)$ によって判断
 - 事後確率が $P(X = A|D) > P(X = B|D)$ であれば、Aが使われた可能性が高い
- 事後確率の計算：
$$P(X|D) = \frac{P(D|X)P(X)}{P(D)}$$

事後確率の計算： ベイズの定理を用いる

- 事後確率の計算には「ベイズの定理」：

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)}$$

事前確率
(今回は1/2)



T. Bayes.

- 判断基準： $P(X = A|D) > P(X = B|D)$
 $\leftrightarrow P(D|X = A)P(X = A) > P(D|X = B)P(X = B)$
- 注意：分母 $P(D) = \sum_X P(D|X)P(X)$ を計算する必要はない

- サイコロのパラメータ $\{p_k^A\}_{k=1}^6, \{p_k^B\}_{k=1}^6$ は最尤推定によって推定

サイコロCの出目回数

- $P(D|X = A) = \prod_{k=1}^6 p_k^A n_k^C > P(D|X = B) = \prod_{k=1}^6 p_k^B n_k^C$ で判断

- では、サイコロAが2個サイコロBが1個あった場合にはどうなる？

ポアソン分布の最尤推定： 標本平均がパラメータの最尤推定量になる

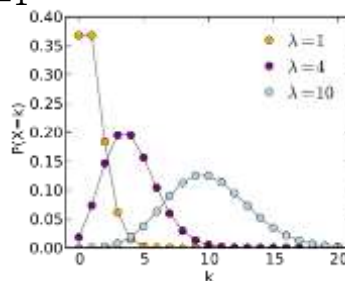
■ ポアソン分布： $P(X = k | \lambda) = \frac{\lambda^k}{k!} \exp(-\lambda)$

■ データ： x_1, x_2, \dots, x_n に対する対数尤度：

$$L(\lambda) = \sum_{i=1}^n \log P(X = x_i | \lambda) = \log \lambda \sum_{i=1}^n x_i - n\lambda + \text{const.}$$

■ パラメータの最尤推定量：

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$



https://en.wikipedia.org/wiki/Poisson_distribution#/media/File:Poisson_pmf.svg

KYOTO UNIVERSITY

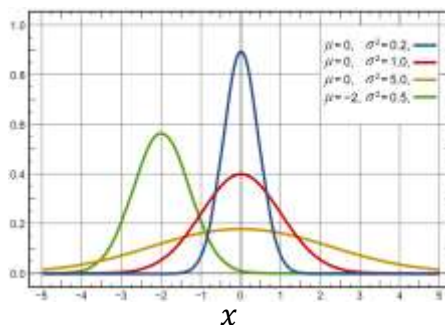
13

練習： 正規分布のパラメータの最尤推定

■ 正規分布： $f(x) = N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

■ パラメータ：平均 μ と分散 σ^2 の最尤推定量を求めてみよう

1. データ： x_1, x_2, \dots, x_n に対する対数尤度をつくる
2. パラメータについての最大化問題を解く



14

KYOTO UNIVERSITY