情報通信システム(特)論II
Information and Communication Systems II

第5回講義資料
Lecture notes 5

ベイズ推定
Bayesian Inference

豊橋技術科学大学
Toyohashi University of Technology
電気・電子情報工学系
Department of Electrical and Electronic Information Engineering
准教授　竹内啓悟
Associate Professor　Keigo Takeuchi

母数の点推定 (Point estimation of parameters)

母集団分布$P(X\,|\theta)$*を既知とする。標本Xから標本値xが得られたときに、母数θを点推定したい。

Point-estimate the parameter from the known population distribution and a realization of the sample.

最尤推定 (Maximum likelihood (ML) estimation)

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, P(X = x\,|\theta)$$

- 尤度$P(X = x\,|\theta)$が最大になる母数$\hat{\theta}$を推定量とする。

  Select an estimator that maximizes the likelihood.

- 多数の無作為標本値を利用できる場合には、最尤推定の性能を上回る推定量は存在しない。

No estimators outperform the ML estimator, when many realizations of random sampling can be utilized.

ベイズ推定(Bayesian inference)

有限個の標本値しか得られない場合に、母数の事前情報を使って推定性能を改善したい。

When a finite number of realizations are given, use a priori information on the parameter to improve the performance in estimation.

* : 母数は確率変数とみなされる。(Regard the parameter as a random variable.)

ベイズ推定における基本仮定 (Basic assumptions on Bayesian inference)

- 母数θは事前分布$P(\theta)$に従って発生する。

  The parameter is drawn from a priori distribution.

- 事前分布は既知である。

  The a priori distribution is known.

ベイズ推定では、データを観測した後に母数が従う分布(事後分布)$P(\theta\,|X)$が使用される。

The a posteriori distribution—distribution of the parameter after observing the data—is used in Bayesian inference.

ベイズの公式 (Bayes' formula)：

$$P(\theta\,|X) = \frac{P(X\,|\theta)P(\theta)}{P(X)} \qquad (5.1)$$

事後分布は、母集団分布$P(X\,|\theta)$と事前分布$P(\theta)$から計算できる。

The a posteriori distribution can be computed with the population distribution and the a priori distribution.

証明：同時分布$P(X,\theta)$を二通りの方法で表現する。

Proof　We represent the joint distribution in two ways.

$$P(X,\theta) = P(\theta\,|X)P(X) = P(X\,|\theta)P(\theta)$$

中辺と右辺を$P(X)$で割ると、ベイズの公式(5.1)を得る。

Dividing the last two expressions by $P(X)$, we arrive at Bayes' formula (5.1). ■

例1：コイン投げ (Example 1: Coin tossing)

標本 (Sample) $X=1$：表 (head)、$X=0$：裏 (tail)

母集団分布

Population distribution

$$P(X=x\,|\theta)=\begin{cases}\theta & \text{for } x=1\\ 1-\theta & \text{for } x=0.\end{cases}$$

母数$\theta\in[0,1]$は表が出る確率を表す。

The parameter represents the probability of the occurrence of a head.

事前確率密度関数(A priori probability density function (pdf)) $p(\theta)=1$

本資料では、$p_\theta(\theta)$を$p(\theta)$と略記する。(In this notes, $p_\theta(\theta)$ is simply written as $p(\theta)$.)

表を1回観測したときの母数θの最尤推定値は、$\hat{\theta}_{\mathrm{ML}}=1$である。

The ML estimate of the parameter is $\hat{\theta}_{\mathrm{ML}}=1$ when one head is observed.

$$\because\quad \hat{\theta}=\operatorname*{argmax}_{\theta\in[0,1]} P(X=1\,|\theta)=\operatorname*{argmax}_{\theta\in[0,1]}\theta=1$$

2番目の等号は母集団分布の定義から従う。

The second equality follows from the definition of the population distribution.

データのみを使って推定すると、確率1で表が出ると推定するしかない。

The estimation based only on the data just predicts that a head occurs with probability 1.

例１：コイン投げ (Example 1: Coin tossing)

事前確率密度関数

A priori pdf

$$p(\theta) = 1$$

母集団分布

Population distribution

$$P(X = x\,|\theta) = \begin{cases} \theta & \text{for } x = 1 \\ 1 - \theta & \text{for } x = 0. \end{cases}$$

表を一回観測した後のθの事後確率密度関数を計算しよう。

Compute the a posteriori pdf of θ given an observation of one head.

最初に、ベイズの公式(5.1)の分母を計算する。

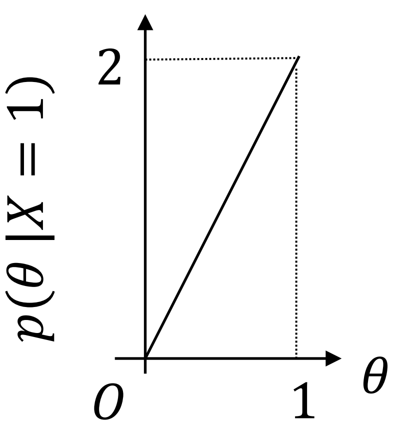
We first compute the denominator in Bayes' formula.

$$P(X = 1) = \int_0^1 P(X = 1\,|\theta)p(\theta)d\theta = \int_0^1 \theta d\theta = \frac{1}{2}$$

この式をベイズの公式に代入すると、次の事後確率密度関数を得る。

Substituting this equation into Bayes' formula, we obtain the following a posteriori pdf:

$$p(\theta\,|X = 1) = \frac{P(X = 1\,|\theta)p(\theta)}{P(X = 1)} = 2\theta$$



この事後確率密度関数からθをどう推定すべきか？

How should we estimate the parameter from this a posteriori pdf?

例2：ビット列の伝送 (Example 2: Transmission of a bit sequence)

標本(Sample) $\boldsymbol{X} = \{X_i \in \{0,1\}: i = 1,2,3\}$

X_iはi番目の受信ビットを表す。(X_i represents the ith received bit.)

母集団分布 Population distribution

$$P(\boldsymbol{X}\,|\boldsymbol{\theta}) = \prod_{i=1}^{3} P(X_i\,|\theta_i),\quad P(X_i = x|\theta_i) = \begin{cases} p & \text{for } x \neq \theta_i \\ 1-p & \text{for } x = \theta_i \end{cases}$$

母数$\boldsymbol{\theta} = \{\theta_i \in \{0,1\}: i = 1,2,3\}$は送信ビット列に対応する。

The parameters correspond to a transmitted bit sequence.

各送信ビットθ_iは独立に確率$p(< 1/2)$でビット反転する。

Each transmitted bit is independently bit-flipped with probability p $(< 1/2)$.

事前分布 a priori distribution

$$P(\boldsymbol{\theta}) = \frac{1}{2} \quad \text{for } \boldsymbol{\theta} \in \{\boldsymbol{\theta}_0 = 000,\ \boldsymbol{\theta}_1 = 111\}$$

ビット列$\boldsymbol{x} = 101$が受信されたときに、送信ビット列を推定しよう。

Estimate the transmitted bit sequence when the bit sequence has been received.

最尤推定値(ML estimate)： $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = 111$

$$\because \quad P(\boldsymbol{X} = \boldsymbol{x}\,|\boldsymbol{\theta} = \boldsymbol{\theta}_0) = p^2(1-p) < p(1-p)^2 = P(\boldsymbol{X} = \boldsymbol{x}\,|\boldsymbol{\theta} = \boldsymbol{\theta}_1)$$

例2：ビット列の伝送 (Example 2: Transmission of a bit sequence)

ビット列$\boldsymbol{x} = 101$が受信されたときの$\boldsymbol{\theta}$の事後分布を計算しよう。

Compute the a posteriori distribution of the parameter when the bit sequence has been received.

最初に、ベイズの公式(5.1)の分母を計算する。

We first compute the denominator in Bayes' formula.

$$P(\boldsymbol{X} = \boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x}\,|\boldsymbol{\theta} = \boldsymbol{\theta}_0)P(\boldsymbol{\theta} = \boldsymbol{\theta}_0) + P(\boldsymbol{X} = \boldsymbol{x}\,|\boldsymbol{\theta}_1)P(\boldsymbol{\theta} = \boldsymbol{\theta}_1)$$

$$= \frac{1}{2}p^2(1-p) + \frac{1}{2}p(1-p)^2$$

(前ページの最後の式を参照)

(See the last equation in the preceding page)

この式をベイズの公式に代入すると、次の事後分布を得る。

Substituting this equation into Bayes' formula, we arrive at the following a posteriori distribution:

$$p(\boldsymbol{\theta}\,|\boldsymbol{X} = \boldsymbol{x}) = \frac{P(\boldsymbol{X} = \boldsymbol{x}\,|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{X} = \boldsymbol{x})} = \begin{cases} p & \text{for } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ 1-p & \text{for } \boldsymbol{\theta} = \boldsymbol{\theta}_1 \end{cases}$$

この事後分布からθをどう推定すべきか？

How should we estimate the parameter from this a posteriori distibution?

評価基準の例1 (Example 1 for evaluation criteria)

平均二乗誤差
Mean-square error (MSE)
$$\mathbb{E}\left[(\theta - \hat{\theta})^2\right]$$

事後平均推定量
A posteriori mean estimator (PME)
$$\hat{\theta}_{\mathrm{PME}} = \mathbb{E}[\theta \mid X]$$

定理5.1 :	事後平均推定量は平均二乗誤差を最小にする。
Theorem 5.1	The PME minimizes the MSE.

事後平均推定量は、**最小平均二乗誤差推定量**とも呼ばれる。

The PME is also called the minimum MSE (MMSE) estimator.

証明(proof) : $\mathbb{E}\left[(\theta - \hat{\theta})^2\right] = \mathbb{E}_X\left\{\mathbb{E}_{\theta|X}\left[\left\{\left(\theta - \hat{\theta}_{\mathrm{PME}}\right) + \left(\hat{\theta}_{\mathrm{PME}} - \hat{\theta}\right)\right\}^2\right]\right\}$

推定量はXの関数なので、条件付き平均$\mathbb{E}_{\theta|X}$の計算では定数とみなせる。

Any estimator is a function of X, and thus regarded as a constant in calculating the conditional expectation.

$\hat{\theta}_{\mathrm{PME}} = \mathbb{E}[\theta \mid X]$なので、平方展開すると、$2\left(\theta - \hat{\theta}_{\mathrm{PME}}\right)\left(\hat{\theta}_{\mathrm{PME}} - \hat{\theta}\right)$は消える。

Since $\hat{\theta}_{\mathrm{PME}} = \mathbb{E}[\theta \mid X]$ holds, the cross term vanishes in expanding the square.

$$\mathbb{E}\left[(\theta - \hat{\theta})^2\right] = \mathbb{E}\left[\left(\theta - \hat{\theta}_{\mathrm{PME}}\right)^2\right] + \mathbb{E}\left[\left(\hat{\theta}_{\mathrm{PME}} - \hat{\theta}\right)^2\right] \geq \mathbb{E}\left[\left(\theta - \hat{\theta}_{\mathrm{PME}}\right)^2\right]$$

等号は$\hat{\theta} = \hat{\theta}_{\mathrm{PME}}$のときに限る。(The equality holds only for $\hat{\theta} = \hat{\theta}_{\mathrm{PME}}$.) ■

例1 : コイン投げ (Example 1: Coin tossing)

表を一回観測した後のθの事後確率密度関数は以下であった。

The a posteriori pdf of θ given an observation of one head has been derived as follows:

$$p(\theta\,|X=1)=2\theta$$

事後平均推定値$\hat{\theta}_{\mathrm{PME}}=\mathbb{E}[\theta\,|X=1]$を計算しよう。

Compute the a posteriori mean estimate.

$$\hat{\theta}_{\mathrm{PME}}=\int_0^1 \theta p(\theta\,|X=1)d\theta=\int_0^1 2\theta^2 d\theta=\frac{2}{3}$$

データと事前情報を使って推定すると、確率2/3で表が出るという推定値を得る。

The estimation based on the data and a priori information predicts that a head occurs with probability 2/3.

比較 : 最尤推定の場合(cf: Case of the ML estimation)

$$\hat{\theta}_{\mathrm{ML}}=1$$

データのみを使って推定すると、確率1で表が出ると推定するしかない。

The estimation based only on the data just predicts that a head occurs with probability 1.

評価基準の例2 (Example 2 for evaluation criteria)

誤り確率:

Error probability

$$p_{\mathrm{er}} = \mathbb{E}\left[1(\theta \neq \hat{\theta})\right], \qquad 1(\mathrm{true}) = 1, \qquad 1(\mathrm{false}) = 0$$

最大事後確率推定量:

Maximum a posteriori (MAP) estimator

$$\hat{\theta}_{\mathrm{MAP}} = \operatorname*{argmax}_{\theta} P(\theta \mid X)$$

定理5.2:	最大事後確率推定量は誤り確率を最小にする。
Theorem 5.2	The MAP estimator minimizes the error probability.

証明(proof): 簡単化のため、$\theta \in \{0, 1\}$を仮定する。(For simplicity, assume $\theta \in \{0, 1\}$.)

$$p_{\mathrm{er}} = \mathbb{E}_X\left\{\mathbb{E}_{\theta|X}\left[1(\theta \neq \hat{\theta})\right]\right\} = \mathbb{E}_X\left\{1\left(\hat{\theta} \neq 0\right)P(\theta = 0|X) + 1\left(\hat{\theta} \neq 1\right)P(\theta = 1|X)\right\}$$

定義により、$1\left(\hat{\theta} \neq 0\right)$と$1\left(\hat{\theta} \neq 1\right)$はどちらか一方が0で、もう一方が1となる。

By definition, either of the two indicator functions takes 0, and the other takes 1.

大きい方の事後確率の係数が0になるように$\hat{\theta}$を定義すると、p_{er}は最小となる。

p_{er} is minimized when one selects $\hat{\theta}$ such that the coefficient of the larger a posteriori probability vanishes.

この推定量は最大事後確率推定量に他ならない。

This estimator is exactly equal to the MAP estimator. ■

例2：ビット列の伝送 (Example 2: Transmission of a bit sequence)

ビット列$\boldsymbol{x} = 101$が受信されたときの$\boldsymbol{\theta}$の事後分布は以下であった。

We have computed the following a posteriori distribution of $\boldsymbol{\theta}$ when the bit sequence has been received:

$$p(\boldsymbol{\theta} \mid \boldsymbol{X} = \boldsymbol{x}) = \begin{cases} p & \text{for } \boldsymbol{\theta} = 000 \\ 1 - p & \text{for } \boldsymbol{\theta} = 111 \end{cases}$$

最大事後確率推定値$\hat{\theta}_{\mathrm{MAP}}$を計算しよう。(Compute the MAP estimate.)

$p < 1/2$なので、(Because of $p < 1/2$,) $\hat{\theta}_{\mathrm{MAP}}$=111

比較：最尤推定の場合(cf: Case of the ML estimation) $\hat{\theta}_{\mathrm{ML}} = 111$

この場合は、最大事後確率推定値と最尤推定値とが等しい。

In this case, the MAP estimate is equal to the ML estimate.

この一致は偶然か？(Does this coincidence occur by chance?)

偶然の一致ではない。(No, it does not.)

定理5.3: Theorem 5.3	事前分布が一様分布のとき、最大事後確率推定量は最尤推定量と一致する。 If the a priori distribution is uniform, the MAP estimator is equivalent to the ML estimator.

証明(proof): ベイズの公式(5.1)を最大事後確率推定の定義に代入すると、

Substituting Bayes' formula into the definition of the MAP estimator yields

$$\hat{\theta}_{\mathrm{MAP}} = \underset{\theta}{\operatorname{argmax}} P(\theta \mid X) = \underset{\theta}{\operatorname{argmax}} \frac{P(X \mid \theta)P(\theta)}{P(X)} = \underset{\theta}{\operatorname{argmax}} P(X \mid \theta)P(\theta) \ .$$

最後の等号は、θに関する最大化に$P(X)$は影響を与えないためである。

The last equality holds because $P(X)$ provides no impacts on the maximization with respect to θ.

事前分布$P(\theta)$ はθに依存しないため、同じ議論を繰り返して、

Since the a priori distribution is independent of θ, we repeat the same argument to obtain

$$\hat{\theta}_{\mathrm{MAP}} = \underset{\theta}{\operatorname{argmax}} P(X \mid \theta) = \hat{\theta}_{\mathrm{ML}} \, .$$

ベイズ推定の基本仮定に対する批判

Criticisms against the basic assumptions in Bayesian inference

- 母数θは事前分布に従って発生する。

  The parameter is drawn from a priori distribution.

- 事前分布は既知である。

  The a priori distribution is known.

批判1 (Criticism 1)：

事前分布など存在するのか？存在したとしても、それを正確に知ることはできるのか？

Is there really the a priori distribution? Even if there were, would it be possible to know the distribution exactly?

- 事後分布は低計算量で計算可能である。

  The a posteriori distribution is computable with low complexity.

批判2(Criticism 2)：

仮に知ることができたとしても、事後分布は計算できるのか？

Even if it were possible, could we compute the a posteriori distribution?

統計学の歴史(History of Statistics)

トーマス・ベイズ(Thomas Bayes) 1702－1761

主観確率(Subjective probability)：ベイズ統計学(Bayesian statistics)の始まり

確率とは、人間の主観的な信念の度合いを数値化したものである。

Probability is a value to quantify human subjective “belief.”

ロナルド・フィッシャー(Ronald A. Fisher) 1890－1962

客観確率(Objective probability)：頻度主義統計学(Frequentist statistics)の始まり

確率とは、実験によってのみ判明する事象の発生頻度である。

Probability is a relative frequency of occurrence determined only via experiments.

彼は前ページの批判をして、統計学から主観を徹底的に排除した。

He provided the criticisms in the preceding page, and eliminated subjectivity from statistics.

統計学の歴史(History of Statistics)

20世紀後半(Latter half of 20th century)〜今日(now)

客観主義ベイズ統計学(Objective Bayesian statistics)

客観的にもっともらしい事前分布の研究ー無情報事前分布

Researches of “objective” a priori distributions—non-informative a priori distirbutions

批判１に対する対抗（Defence against criticism 1）

計算機科学の進化(Evolution of computer science)

事後分布を高速で近似計算する手法の発展

Development of methods for computing the a posteriori distribution quickly and approximately.

批判2に対する対抗（Defence against criticism 2）

ベイズ統計学は今日の情報通信技術を支える基盤である。

Bayesian statistics is a base that supports modern Information and communication technology.