ベイズ統計の理論と方法

渡辺　澄夫 著

コ ロ ナ 社

ま　え　が　き

本書の目的は，ベイズ統計の理論と方法について基礎的な事柄を解説することである。

ベイズ推測は統計学において重要な方法の一つであり広く応用され優れた実用性をもっているにもかかわらず，大学あるいは大学院において必ずしも十分に詳しい講義が行われてはいないように思われる。このため，実務の中で必要になってから生じる種々の疑問について，それが基礎的なものであればあるほど問いかけるあてもないという状況ではないだろうか。

そこで本書ではベイズ推測に関して，あまりにも当然過ぎるために普段は説明がなされていないことや，多くの人が疑問に思いながらも通り過ぎてしまうことについて解説を行う。また，やさしい計算であるからという理由でその前提が成り立たない場合にも誤って応用されてきた理論の限界を述べ，反対に数学教室以外では習わない数理が必要になるという理由で実用上でも大切であるにもかかわらず知られていなかった一般的な法則を明らかにする。

ベイズ推測を用いて構造をもつ推論システムを解析し構築するという課題は今日の科学と技術の中でますます重要度を増しているが，その発展が実り多いものとなるためには，高度化し複雑化していくものを支えることができる確かな基盤が必要である。大きな樹には大きな根の広がりが不可欠である。学問が発展すればするほど「確かに拠って立つことができる場所」の上にそれが築かれていなくてはならないからである。

本書を読むのに必要となる予備知識は，大学初年度に習う線形代数と微分積分だけで十分であり，初等確率論をまだ学んでいない読者は最後の章を確認しながら読み進めていただきたい。また，ベイズ統計の理論と方法をつくるためには線形代数・微分積分・初等確率論には含まれていない基礎数学も必要にな

るが，そのような場所においては重要な概念について初等的に理解できるように導入部分を加えている。自然科学・人文社会科学・情報科学の課題に挑む読者にとって本書が基礎となることを願っている。

2012 年 2 月

渡辺 澄夫

目　　　次

1.　は　じ　め　に

2.　基　礎　概　念

3.　正　則　理　論

4.　一 般 理 論

5.　事後分布の実現

6.　ベイズ統計学の諸問題

7.　ベイズ統計の基礎

8.　初等確率論の基礎

1 はじめに

1 章では，ベイズ推測を定義し，本書で述べることのあらましを述べる。

ベイズ推測についてすでに学んでいて多くの疑問や問いかけを抱いている読者も，とりあえず 1 章では本書におけるベイズ推測の定義を確認し，本書の全体の構造を眺めていただきたい。

本書で必要になる初等確率論については 8 章にまとめてある。初等確率論，例えば「条件付き確率」や「カルバック・ライブラ情報量」などについて初めて出会った読者は本書を読み始める前に，8 章を確認していただきたい。なお，自然科学や科学技術の領域では「確率分布」という言葉で確率分布だけでなく確率密度関数をも表すことが多いので，本書でも同じ定義を用いることにする。

1.1 ベイズ推測の定義

まずベイズ推測を定義する。

N と n を自然数とする。N 次元実ユークリッド空間 $\mathbb{R}^N$ 上に n 個の点の集合 $x_1, x_2, ..., x_n$ がある場合を考える。

$$x_i \in \mathbb{R}^N \qquad (i = 1, 2, 3, ..., n) \tag{1.1}$$

である。これらの n 個の点の集合を**サンプル**と呼ぶ。n 個のサンプルを一つの記号で表すとき

$$x^n = (x_1, x_2, ..., x_n) \tag{1.2}$$

という記号を用いる。すなわち

$$x^n = (x_1, x_2, ..., x_n) \in \mathbb{R}^N \times \mathbb{R}^N \times \cdots \times \mathbb{R}^N \tag{1.3}$$

である。サンプル $x_1, x_2, ..., x_n$ が，ある確率分布 $q(x)$ に独立に従う確率変数の実現値であるとしよう。すなわち x^n を，$(\mathbb{R}^N)^n$ 上の分布

$$q(x^n) = \prod_{i=1}^{n} q(x_i) = q(x_1)q(x_2)\cdots q(x_n) \tag{1.4}$$

をもつ確率変数

$$X^n = (X_1, X_2, ..., X_n)$$

の実現値であると考える。この $q(x)$ を**真の分布**と呼ぶ。現実の問題では，真の分布は不明であって，サンプルだけが与えられることが多い。与えられたサンプルから真の分布を推測することを**統計的推測**あるいは**統計的学習**という。統計的推測によって得られる結果はサンプルに依存して確率的に変動するから，その性質を解明するためには確率的な挙動を調べなくてはならない。そこで以下では，サンプルを確率変数であると考え，確率的に変動するものに対する統計的推測を定義する。

注意 1 現実の問題においては数値データとして x^n が与えられるだけであり，確率変数 X^n が与えられるのではない。もしもサンプルを確率変数の実現値であると考えなければ，確率的な推論を行うことはできない。しかしながらサンプルを確率変数の実現値だと考えることにより，統計的推測において成り立つ一般的な法則を導出することができ，その法則の中にワンセットの実現値 x^n を捉（とら）えることにより構造的な視点からサンプルを考察することができるようになる。

注意 2 サンプルから真の分布を推測することを統計学では統計的推測といい，情報科学では統計的学習という。この二つは歴史的な起源は必ずしも同じではないが，今日では両者はまったく同一のものとなっている。

サンプルを表す確率変数を X^n とする。その関数

$$f(X^n) = f(X_1, X_2, ..., X_n)$$

が与えられたとき，平均値をとる操作 $\mathbb{E}[\ ]$ を

$$\mathbb{E}[f(X^n)] = \int\int\cdots\int f(x_1, x_2, ..., x_n)\prod_{i=1}^{n} q(x_i)dx_i$$

と表記する。この平均値 $\mathbb{E}[\ ]$ を「サンプルの現れ方に対する平均値」と呼ぶ。統計的推測においては，サンプル X^n を用いて推測を行った後，同じ真の分布 $q(x)$ から，サンプル X^n とは独立な確率変数 X を発生させて推測結果のよさを評価したいことがある。この確率変数 X の関数 $f(X)$ についての平均を

$$\mathbb{E}_X[f(X)] = \int f(x)q(x)dx$$

と表記する。

さて，真の分布を推測するときに人間が準備するものを考える。ベイズ推測においては，パラメータ $w \in W \subset \mathbb{R}^d$ が与えられたときの $x \in \mathbb{R}^N$ の上の条件付き確率分布 $p(x|w)$ と $w \in W$ 上の確率分布 $\varphi(w)$ とが必要である。このとき $p(x|w)$ を**確率モデル**といい，$\varphi(w)$ を**事前分布**という。

統計的推測が行われる一般的な状況では，真の分布 $q(x)$ は不明であるから，確率モデル $p(x|w)$ と事前分布 $\varphi(w)$ が真の分布の推測において適切であるかどうかは不明であることが多い。サンプルだけが与えられたとき，真の分布が不明であるにもかかわらず，確率モデルと事前分布をどのように設計したらよいか，設計された確率モデルと事前分布はどの程度に適切であるといえるか，という課題は重要な課題であるが，そうしたことについては，順を追って考察していくことにし，ここでは，ひとまず任意の三つ組 $(q(x), p(x|w), \varphi(w))$ が与えられていると考えることにする。すなわち，確率モデルは真の分布に対して適切でなくてもよいし，事前分布も事前の確率を表しているものでなくてもよい。

定数 β を $0 < \beta < \infty$ 満たす実数とし，これを**逆温度**と呼ぶ。パラメータ w の逆温度 β の**事後分布**を

$$p(w|X^n) = \frac{1}{Z_n(\beta)}\varphi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta} \tag{1.5}$$

と定義する。ここで $Z_n(\beta)$ は事後分布の $w \in W$ に関する積分が 1 になるように定めた定数である。すなわち

$$Z_n(\beta) = \int_W \varphi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta} dw \tag{1.6}$$

である。この値 $Z_n(\beta)$ を**分配関数**という。$\beta = 1$ のとき，$Z_n(1)$ を**周辺尤度**という。分配関数はサンプル X^n に依存して確率的に変動するから確率変数である。

注意 3「条件付き確率分布」という概念に初めて出会った読者は，8.3 節を確認していただきたい。もしも事前分布がパラメータの本当の事前確率であり，確率モデルがパラメータが与えられたときの X の本当の条件付き確率分布を表しているときには，$\beta = 1$ のときの事後分布はサンプルが観測されたという下でのパラメータの本当の条件付き確率分布である。しかしながら，本書においては，確率モデル $p(x|w)$ は真の分布に対して適切であることを仮定しないし，事前分布 $\varphi(w)$ もパラメータの本当の事前の確率であることを仮定しない。したがって事後分布も，サンプルが得られたときのパラメータの本当の条件付き確率分布とはかぎらない。ここでは事後分布というものを上記のように定義したのであって，真の事後分布を導出したのではないことに注意していただきたい。こうした状況では，事後分布は人間が定めたものであるにすぎず，無条件に正しいと信じられるものではないのである。

注意 4 本書では，事後分布は一般の逆温度 β を用いて定義する。ベイズ統計学においては，$\beta = 1$ の場合が特別に重要であり，通常の本や論文で「事後分布」あるいは「ベイズ推測」という言葉が用いられる場合には $\beta = 1$ の場合だけを意味していることが多い。本書においても $\beta = 1$ の場合が最も重要である。しかしながら，これから考察していくように，一般の $\beta \neq 1$ の場合に生じる現象もベイズ統計の理論と方法を考えるうえで大切であるので，本書では一般の

β の場合も考える。β についての制限が書いてない場合には一般の β を想定しているという意味であり，$\beta = 1$ の場合だけを考えているときには「$\beta = 1$ の場合は」という条件を明記する。

パラメータ w の関数 $f(w)$ が与えられたとき，事後分布 $p(w|X^n)$ による平均を

$$\mathbb{E}_w[f(w)] = \int f(w)p(w|X^n)dw \tag{1.7}$$

と表記する。すなわち $\mathbb{E}_w[\ ]$ は事後分布に関する平均値を求める操作を表すこととする。平均値 $\mathbb{E}_w[f(w)]$ はパラメータについての積分を計算しているのでパラメータの関数ではない。しかしながら，$\mathbb{E}_w[\ ]$ は，サンプル X^n の値に依存する事後分布で平均を行うことを表しているのであって，平均値 $\mathbb{E}_w[f(w)]$ はサンプル X^n の変動に伴って変動する。すなわち $\mathbb{E}_w[f(w)]$ は確率変数である。

事後分布によって確率モデル $p(x|w)$ を平均したもの

$$p^*(x) = \mathbb{E}_w[p(x|w)] = \int p(x|w)p(w|X^n)dw \tag{1.8}$$

を**予測分布**という。これを

$$p^*(x) = p(x|X^n) \tag{1.9}$$

と書く場合もある。ベイズ推測とは

「真の確率分布 $q(x)$ は，おおよそ $p^*(x)$ であろう」

と推測することである。予測分布 $p^*(x)$ もまた，サンプル X^n に応じて確率的に変動する。

確率モデルも事前分布も真の分布を知らない人間が定めたものであるから，予測分布 $p^*(x)$ もまた人間が定めたものである。これは，あくまでも人間の推測であるうえにサンプルの確率的な変動もあるのであるから，真の分布 $q(x)$ と

ぴったりと一致することはほとんどないであろう。しかしながら，$q(x)$ に対して組 (p,φ) が適切であり，サンプルの個数 n が多ければ，この推測はある程度よい推測ではないかと期待される。もしも組 (p,φ) が適切でなければ，予測分布は真の分布から大きくずれているであろう。それでは，予測分布は，どのような条件の下で，どのくらいよい推定になっているのだろうか。真の分布が不明であるにもかかわらず，サンプルだけから予測分布の推測の精度を知ることができるのだろうか。

本書ではベイズ推測について，つぎの問題を考えていく。

1. 予測分布 $p^*(x)$ は，真の分布 $q(x)$ とどのくらい似ているか。
2. できるだけ推測 $p^*(x)$ が真の分布 $q(x)$ と似ているようにするためには，どうするべきか。
3. これらの問題を考えるために拠って立つことができる数学的構造をつくりたい。

注意 5 現実の問題ではサンプル X^n の実現値が偶然の結果として与えられるだけであり，真の分布はわからない。したがって，人間が用意した確率モデルと事前分布が真の分布に対して適切であるかどうかはわからない。これより，推測の結果として得られる予測分布も，真の分布をよく推測しているかどうかわからない。現実の世界の中の統計的推測とは「わからないことを前提として推論を行い，わからない結論に到達する」という

$$\text{わからない} \to \text{推論} \to \text{わからない} \to \cdots \tag{1.10}$$

の繰返しにすぎないのではないか，と感じる人もあることだろう。

そこには信頼するに足りる基盤があるのだろうか。

本書では，つぎのことを述べていく。統計的推測においては，三組（真の分布・確率モデル・事前分布）に依存しない数学的な法則が存在する。この法則は，確率モデルと事前分布が真の分布に対して適切であってもなくても成立するものであり，事後分布が正規分布で近似できてもできなくても成立するものである。この法則のうえに立つことにより，真の分布についての前提を設定し

索　　　　引

—— 著 者 略 歴 ——

1982 年　東京大学理学部物理学科卒業
2001 年　東京工業大学教授
　　　　　現在に至る

ベイズ統計の理論と方法
Theory and Method of Bayes Statistics

2012 年 4 月 12 日　初版第 1 刷発行　★

検印省略

著　　者　渡（わた）　辺（なべ）　澄（すみ）　夫（お）
発 行 者　株式会社　コ ロ ナ 社
　　　　　代 表 者　牛 来 真 也
印 刷 所　三 美 印 刷 株 式 会 社

112–0011　東京都文京区千石 4–46–10
発行所　株式会社　コ　ロ　ナ　社
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替 00140–8–14844・電話(03)3941–3131(代)
ホームページ　http://www.coronasha.co.jp

ISBN 978–4–339–02462–3　（金）　（製本：愛千製本所）
Printed in Japan