

# 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必要である。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

ii ま え が き

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

1.1	ベイズ推測の定義	1
1.2	考察される量	7
1.2.1	分配関数と自由エネルギー	7
1.2.2	推測と汎化	9
1.2.3	計算できる例	11
1.3	さまざまな推測方法	16
1.4	事後分布の例	18
1.5	確率モデルの例	22
1.5.1	確率モデルがわかっている場合	22
1.5.2	確率モデルが仮のものである場合	23
1.6	本書の概略	24
1.7	一般的注意	25
1.7.1	本書の厳密性について	25
1.7.2	表記法	26
1.8	質問と回答	28
	章末問題	29

## 2. 基礎概念

2.1	真の分布と確率モデルの関係	30
-----	---------------	----

2.2 理論の基礎	40
2.2.1 基礎概念	40
2.2.2 正規化された変量	41
2.2.3 キュムラントと母関数	44
2.3 ベイズ統計理論の構造	49
2.4 質問と回答	50
章末問題	51

### 3. 正則理論

3.1 基礎数学の公式	53
3.1.1 転置行列, トレース, 行列式	53
3.1.2 対称行列, 固有値, 正定値行列	55
3.1.3 積分公式	56
3.1.4 平均値の定理	57
3.2 分配関数の挙動	58
3.2.1 準備	59
3.2.2 分配関数の非主要項	61
3.2.3 分配関数の主要項	62
3.3 スケーリング	67
3.4 汎化損失と経験損失	69
3.5 事後確率最大化法	72
3.5.1 推定量の漸近分布	72
3.5.2 汎化誤差と経験誤差	75
3.6 サンプルから計算する方法	76
3.6.1 自由エネルギー	77
3.6.2 汎化損失と経験損失	78

3.7 質問と回答	84
章末問題	86

## 4. 一般理論

4.1 多様体	88
4.2 標準形	91
4.2.1 特異点解消定理	91
4.2.2 標準形	97
4.3 状態密度の挙動	100
4.3.1 超関数	100
4.3.2 状態密度関数	103
4.4 統計的推測の一般理論	110
4.4.1 分配関数	110
4.4.2 繰り込まれた事後分布	112
4.5 相転移	121
4.6 事後確率最大化法	125
4.6.1 平均プラグイン法	125
4.6.2 事後確率最大化法	125
4.7 質問と回答	131
章末問題	134

## 5. 事後分布の実現

5.1 マルコフ連鎖モンテカルロ法	135
5.1.1 メトロポリス法	137
5.1.2 ギブス・サンプリング	141

5.1.3 ランジューバン方程式を用いる方法	143
5.1.4 自由エネルギーの近似	145
5.2 平均場近似	148
5.2.1 平均場近似とは	148
5.2.2 変分ベイズ法	153
5.3 質問と回答	161
章末問題	162

## 6. ベイズ統計学の諸問題

6.1 回帰問題	163
6.2 モデルの評価	168
6.2.1 評価の規準	168
6.2.2 バイアスとバリエーション	169
6.2.3 偏差情報量規準	171
6.3 クロスバリデーション	176
6.4 統計的検定	182
6.4.1 ベイズ検定	182
6.4.2 ベイズ検定の例	186
6.5 質問と回答	189
章末問題	191

## 7. ベイズ統計の基礎

7.1 確率モデルと事前分布がわかっているとき	192
7.2 確率モデルあるいは事前分布がわかっていないとき	194
7.3 確率モデルと事前分布	198

7.3.1	指数型分布について	198
7.3.2	線形回帰モデル	199
7.3.3	構造をもつ確率モデル	199
7.3.4	ハイパーパラメータの最適化	200
7.4	質問と回答	201
章末問題		202

## 8. 初等確率論の基礎

8.1	確率分布と確率変数	203
8.2	平均と分散	205
8.3	同時分布と条件付き確率	206
8.4	カルバック・ライブラ情報量	208
8.5	極限定理	209
8.5.1	確率変数の収束	209
8.5.2	大数の法則と中心極限定理	209
8.5.3	経験過程	210
引用・参考文献		213
章末問題解答		215
索引		224

# 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  $x_1, x_2, \dots, x_n$  がある場合を考える。

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

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

$$x^n = (x_1, x_2, \dots, x_n) \quad (1.2)$$

という記号を用いる。すなわち

## 2 1. はじめに

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

である。サンプル  $x_1, x_2, \dots, x_n$  が<sup>s</sup>, 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

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

をもつ確率変数

$$X^n = (X_1, X_2, \dots, X_n)$$

の実現値であると考え。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

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

が与えられたとき、平均値をとる操作  $\mathbb{E}[\ ]$  を

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

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  $\beta$  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  $\beta$  の事後分布を

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

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

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

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<sup>ゆうど</sup>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  $\beta$  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

$\beta$  の場合も考える。 $\beta$  についての制限が書いてない場合には一般の  $\beta$  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

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

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

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

を予測分布という。これを

$$p^*(x) = p(x|X^n) \quad (1.9)$$

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  $(p, \varphi)$  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  $(p, \varphi)$  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はわからない。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

$$\text{わからない} \rightarrow \text{推論} \rightarrow \text{わからない} \rightarrow \dots \quad (1.10)$$

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

	<b>【あ】</b>				
赤池情報量規準	80	ギブス・サンプリング法	141	<b>【さ】</b>	
		ギブス推測	86	最強力検定	184
		ギブスの変分原理	149	最尤推測	16
<b>【い】</b>		逆温度	3	最尤推定量	16
位相空間	89	共役な事前分布	12	座標近傍系	89
一次転移	121	行列式	54	サポート	103
				サンプル	1
		<b>【く】</b>		<b>【し】</b>	
<b>【え】</b>		繰り込まれた事後分布	113	ジェフーズの事前分布	169
エントロピー	8	クロスバリデーション	176	次元	89
エントロピー障壁の問題	139	クロスバリデーション損失	177	事後確率最大化推測	16
				事後確率最大化推定量	16
		<b>【け】</b>		事後微小積分	67, 88
<b>【か】</b>		経験過程	211	事後分布	3
回帰関数	207	経験誤差	42	自己無矛盾条件	151
——のベイズ推測	164	経験誤差関数	41	指数型分布	12
解析多様体	89	——の標準形	99	事前分布	3
カイ二乗分布	83	経験損失	9, 40	実現可能	30
可解モデル	12	——のキウムラント母		実質的にユニーク	33
過学習	131	関数	44	実対数閾値	103, 108
可逆	55	経験対数損失関数	40	自由エネルギー	8, 40
確率過程	59	経験二乗損失	166	周辺確率分布	206
確率収束	209	検出力	183	周辺尤度	4
確率分布	203			縮小ランク回帰モデル	119
確率変数	203	<b>【こ】</b>		条件付き確率分布	206
確率密度関数	203	固有値	55	詳細釣り合い条件	136
確率モデル	3	固有ベクトル	55	状態密度	100
隠れ変数	155	コルモゴロフの拡張定理	211	——の漸近挙動	104
カルバック・ライブラ				真の分布	2
情報量	208	混合指数型分布	154	——に対して最適な	
				パラメータの集合	32
<b>【き】</b>					
棄却域	183				

<b>【す】</b>	
推定量の一致性	72
スケーリング関係	113
スケーリング則	113
<b>【せ】</b>	
正規確率過程	210
正規化された自由エネルギー	42
正規化された分配関数	42
正規交差特異点	92
正規分布	56, 204
正則	32, 55
——の情報量規準	79
潜在変数	155
<b>【そ】</b>	
相対エントロピー	208
相対的に有限な分散	35
相転移	121
相転移点	121
双有理不変量	108
<b>【た】</b>	
台	103
対角行列	55
対称行列	55
対数周辺尤度	8
大数の法則	209
対数尤度比関数	35
多重指数	57
多重度	103
多様体	89
<b>【ち】</b>	
中心極限定理	210
直交行列	55
<b>【て】</b>	
ディリクレ分布	153
デルタ関数	204
転置行列	53

<b>【と】</b>	
統計的学習	2
統計的推測	2
同時確率分布	206
特異点解消定理	91
特異点の解消	91
特異ゆらぎ	117
独立	206
凸集合	205
トレース	53
<b>【に】</b>	
二次転移	121
<b>【ね】</b>	
ネイマン・ピアソンの補題	185
<b>【は】</b>	
バーンイン	138
バイアス	169
ハイパーパラメータ	12
ハイブリッド・モンテカルロ法	140
ハウスドルフ空間	89
白色雑音	144
バリエーション	169
パワー	183
汎化誤差	42
汎化損失	9, 40
——のキュムラント母関数	44
汎化二乗誤差	164
汎化二乗損失	166
汎関数分散	114, 117
<b>【ひ】</b>	
評価の双対性	161
広く使える情報量規準	118
広中の定理	91

<b>【ふ】</b>	
フォッカー・プランク方程式	144
不良設定問題	194
分散共分散	205
分配関数	4
——の主要項	59
——の非主要項	59
<b>【へ】</b>	
平均	205
平均誤差関数	41
平均対数損失関数	31, 40
平均値の定理	58
平均場近似	150
平均場自由エネルギー	150
平均プラグイン推測	16
ベイズ情報量規準	78
ベイズ統計学の状態方程式	118
ベイズ統計の基礎定理	47
ベイズの定理	206
ヘルダーの不等式	205
偏差情報量規準	171
変分自由エネルギー	153
変分ベイズ法	16, 153, 159
<b>【ほ】</b>	
法則収束	209
ポテンシャル障壁の問題	139
<b>【ま】</b>	
マルコフ過程	136
<b>【め】</b>	
メトロポリス法	137
メリン変換	104
<b>【ゆ】</b>	
有意水準	183
尤度関数	16

	<b>【よ】</b>		<b>【り】</b>	レベル	183
予測分布	5		リーブ・フロッグ法	141	
	<b>【ら】</b>		<b>【れ】</b>		
ランジュバン方程式	144		レプリカ交換法	147	

	<b>【A】</b>		<b>【R】</b>		
AIC	80, 82		RIC	53	
	<b>【B】</b>		<b>【T】</b>		
BIC	78, 82		TIC	79	
	<b>【D】</b>		<b>【W】</b>		
DIC	171		WAIC	118	
					<b>【数字】</b>
					1 の分割
					96

— 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

© Sumio Watanabe 2012

2012年4月12日 初版第1刷発行

★

検印省略

著者 わた なべ すみ お  
渡 辺 澄 夫  
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



本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めておりません。

落丁・乱丁本はお取替えいたします