ま　え　が　き

情報技術にとって 1990 年代から 2010 年まではまさしくインフラの時代であったといえる．インターネットの通信インフラの確立，Web データの入力，またデータの格納方法などが最も重要な研究課題であった．その後，インターネット上に大量にデータが蓄積された結果，現在，溢れかえるほどの膨大なデータがインターネット上に氾濫している．いまでは，この大量データをどのように有効活用できるかが最も重要な技術課題となっている．多くの情報技術関連の企業にとって，データ有効活用に競争力をもっていることが重要になってくるであろう．

このような時代に，最もふさわしい技術がベイジアンネットワークである．ベイジアンネットワークは，離散確率分布の近似モデルであり，現在あるモデルの中でも最も仮定が少なく自然なモデルとして予測精度が高いことが知られている．具体的には，因果モデルをグラフ構造としてデータより学習し，その因果モデルを用いてさまざまな推論を行う人工知能の最先端技術である．ベイズ統計を基調としており，単純な推論，予測だけでなく，意思決定理論と組み合わせることにより，より複雑な問題解決を実現できるツールとなる．

また，ベイジアンネットワークは，さまざまな技術を組み合わせて実現されており，モデルとしても最上位にある．そのため，ベイジアンネットワークを深く学習するだけで他の下位モデルや手法を同時に学習したことにもなる．しかし，このことは逆にベイジアンネットワークを学習することが非常に難しいことも示唆している．修士課程から学習を始めても 2 年間で基礎を習得するのは非常に難しい．だからこそ，この技術をもっていることがその人の競争力を高めるのだが，ハードルが高すぎるのはよいことではない．ベイジアンネットワーク研究の最高峰は，UAI (The Conference on Uncertainty in Artificial Intelligence)

という国際会議であるが，ベイジアンネットワークの研究で採録されることは非常に難しくなってきている．トップレベルなので当たり前だというだけでなく，基礎から最先端までの研究を学ぶことがベイジアンネットワークでは非常に難しいことがその理由になっているのかもしれない．

そこで，簡単に早くベイジアンネットワークの基礎から最先端までの研究を学べる書籍を書くことになった．本書は，限られたスペースで，基礎から現在の推論研究と構造学習研究の最先端にたどり着けるように意図して書いた教科書である．また，ページ数の制限とアルゴリズムの理解を中心的に学ばせるために，本書では重要なものを除いてほとんどの定理には証明を掲載していない．また，本書では 2013 年 3 月時点での最先端研究までを掲載しているが，それらに結び付かないトピックはすべて割愛している．しかし，将来的には，それらの流れは大胆に変化していく可能性がある．本書を読んだ後，2013 年以降の UAI の論文を読んでいただくことで最先端の研究に追いつけることを期待している．これからベイジアンネットワークを専門的に学ぼうとする初学者にぜひ役に立てていただきたい．

執筆にあたり，5 章では修士課程 2 年生の小竹遼弥君に，6 章は博士課程 2 年生の宇都雅輝君に，7 章は博士課程 3 年生の森下民平君に多大なる協力を得た．ここに感謝する．

2013 年 3 月

植野　真臣

目　　　　次

1.　確率とビリーフ

2.　ベイジアンネットワークのためのグラフ理論

3.　ベイジアンネットワーク

4.　ジョインツリーアルゴリズムによる推論

5.　ベイジアンネットワークの近似推論

6.　ベイジアンネットワークの学習

7.　条件付き独立性検定による構造学習

1 確率とビリーフ

ベイジアンネットワークはベイズ統計学を基礎としている．本章では，ベイジアンネットワークを学ぶために必要なベイズ統計学の基礎知識を学ぶ．

1.1 確　　　率

本節では，まず，確率（probability）を定義する．確率を定義するためには，それが対象とする事象（event）の集合を定義しなければならない．

定義 1 ($\boldsymbol{\sigma}$ 集合体)　Ω を標本空間（sample space）とし，$\mathcal{A}$ が以下の条件を満たすならば **$\boldsymbol{\sigma}$ 集合体**（σ–field）と呼ぶ．

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$（ただし，$A^c = \Omega \setminus A$）
3. $A_1, A_2, \cdots \in \mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

つまり，たがいに素な事象の和集合により新しい事象を生み出すことができ，それらすべての事象を含んだ集合を σ 集合体と呼ぶ．

σ 集合体上で確率（probability）は以下のように定義される．

定義 2 (確率測度)　いま，σ 集合体 $\mathcal{A}$ 上で，つぎの条件を満たす測度（measure）P を，**確率測度**（probability measure）と呼ぶ（Kolmogorov 1933）．

1. $A \in \mathcal{A}$ について，$0 \leqq P(A) \leqq 1$
2. $P(\Omega) = 1$
3. たがいに素な事象列 $\{A_n\}_{n=1}^{\infty}$ に対して，

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

上の定義では，1. は確率が 0 から 1 の値をとること，2. は全事象の確率が 1 になること，3. はたがいに素な事象の確率はそれぞれの確率の和で求められること，が示されている．特に 3. の条件を確率の加法性（additivity）と呼ぶ．また，三つ組 $(\Omega, \mathcal{A}, P)$ を確率空間（probability space）と呼ぶ．

以下，確率の重要な性質を導こう．

定義 1. の 2. より，$P(A^c) = P(\Omega \setminus A) = P(\Omega) - P(A) = 1 - P(A)$ となることがわかる．これより，以下の定理が成り立つ．

定理 1 (**余事象の確率**)　事象 A の**余事象**（complementary event）の確率は以下のとおりである．

$$P(A^c) = 1 - P(A)$$

また，定義 1. の 2. より，$P(\phi) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$ となる．これより，以下の定理が成り立つ．

定理 2 (**境　界**)

$$P(\phi) = 0$$

さらに，$A \subset B$ のとき，$P(A)$ と $P(B)$ の関係について考えよう．

$A \subset B$ より，$\exists B' : B = A \cup B'$，$A \cap B' = \phi$ が成り立つ．定義 1. の 3. より $P(B) = P(A) + P(B')$，定義 1. の 2. より $0 \leqq P(B') \leqq 1$ が成り立ち，結果として $P(A) \leqq P(B)$ が成り立つことがわかる．これを以下のように**単調性**（monotonicity）と呼ぶ．

定理 3 (**単調性**)　$A \subset B$ のとき $P(A) \leqq P(B)$

定義 2. の 3. では，たがいに素な事象の和の確率はそれぞれの事象の確率の和で求めることができた．では，たがいに素ではない事象の和はどのように求められるのであろうか？　以下のように求められる．

$A \cup B = (A \cap B^c) \cup (A^c \cap B) \cup (A \cap B)$ より, $P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)$ が成り立つ．これより, $P(A) + P(B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)$, そして $P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$ が成り立つ．したがって，$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ となる．これを以下のように**確率の和法則**（additional low of probability）と呼ぶ．

定理 4 (確率の和法則)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

1.2 主 観 確 率

前節で確率を数学的に定義した．しかし，確率の実際的な解釈には二つの立場がある．最も一般的な解釈が，ラプラスの**頻度主義**である．

コインを何百回も投げて表が出た回数（頻度）を数えて，その割合を求めることを考えよう．いま，投げる回数を n とし，表の出た回数を n_1 とすると，$n \to \infty$ のとき，

$$\frac{n_1}{n} \to \frac{1}{2}$$

となることが予想される．このように，何回も実験を繰り返して n 回中，事象 A が n_1 回出たとき，n_1/n を A の確率と解釈するのが頻度主義である．

しかし，この定義では真の確率は無限回実験をしなければならないので得ることは不可能である．また，科学的実験が可能な場合にのみ確率が定義され，実際の人間が扱う不確かさに比べてきわめて限定的になってしまう．

一方, より広く確率を捉（とら）える立場として, 人間の個人的な**主観確率** (subjective probability）として解釈する立場がある．

ベイジアン（Bayesian，ベイズ主義者）は，確率を主観確率として扱う．次節で導出されるベイズの定理を用いる人々をベイジアンだと誤解されているが,

ベイズの定理は確率の基本定理で数学的に議論の余地のないものであり，頻度主義者も用いる．

例えば，松原（2010）では以下のような主観確率の例が挙げられている．

1. 第三次世界大戦が 20XX 年までに起こる確率が 0.01
2. 明日，会社の株式の価格が上がる確率が 0.35
3. 来年の今日，東京で雨が降る確率が 0.5

ベイズ統計では，これらの主観確率は個人の意思決定のための信念として定義され，**ビリーフ**（belief）と呼ばれる．当然，頻度論的確率を主観確率の一種とみなすことができるが，その逆は成り立たない．本書では，ベイズ統計の立場に立ち，確率をビリーフの立場で解釈する．ビリーフの具体的な決定の仕方など厳密な理論に興味のある読者は Bernardo and Smith（1994），Berger（1985）を参照されたい．

1.3　条件付き確率と独立

本節では，条件付き確率と独立を定義する．

定義 3 (条件付き確率)　$A \in \mathcal{A}$，$B \in \mathcal{A}$ について，事象 B が起こったという条件の下で，事象 A が起こる確率を**条件付き確率**（conditional probablity）と呼び，

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

で示す．

このとき，$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ より以下の乗法公式が成り立つ．

定理 5 (乗法公式)

$$P(A \cap B) = P(A|B)P(B)$$

このとき，$P(A \cap B)$ を A と B の**同時確率**（joint probability）と呼ぶ．

つぎに，事象の独立を以下のように定義する．

定義 4 (**独　立**)　ある事象の生起する確率が，他のある事象が生起する確率に依存しないとき，二つの事象は**独立**（independent）であるという．すなわち事象 A と事象 B が独立とは $P(A|B) = P(A)$ であり，

$$P(A \cap B) = P(A)P(B)$$

が成り立つことをいう．

さらに乗法公式を一般化すると以下のチェーンルールが導かれる．

$$P(A \cap B \cap C) = P(A|B, C)P(B \mid C)P(C)$$

これは，3 個以上の事象にも拡張できるので，**チェーンルール**（chain rule）は以下のように書ける．

定理 6 チェーンルール N 個の事象 $\{A_1, A_2, \cdots, A_N\}$ について

$$\begin{aligned}&P(A_1 \cap A_2 \cap \cdots \cap A_N)\\&= P(A_1|A_2, A_3, \cdots, A_N)P(A_2 \mid A_3, A_4, \cdots, A_N) \cdots P(A_N)\end{aligned}$$

が成り立つ．

1.4　ベイズの定理

本節では，条件付き確率より，ベイズ統計にとって最も重要なベイズの定理を導出する．

ベイズの定理を導出する前に，たがいに背反な事象 $A_1, A_2, \cdots, A_n$, $(A_i \in \mathcal{A})$ が全事象 Ω を分割しているとき，事象 $B \in \mathcal{A}$ について以下が成り立つことがわかる．

$$\sum_{i=1}^{n} P(A_i)P(B|A_i) = \sum_{i=1}^{n} P(A_i)\frac{P(A_i \cap B)}{P(A_i)}$$
$$= \sum_{i=1}^{n} P(A_i \cap B) = P(\Omega \cap B) = P(B)$$

これを以下の全確率の定理と呼ぶ.

定理 7 (全確率の定理（total probability theorem))　たがいに背反な事象 $A_1, A_2, \cdots, A_n$,（$A_i \in \mathcal{A}$) が全事象 Ω を分割しているとき，事象 $B \in \mathcal{A}$ について，$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$ が成り立つ.

全確率の定理より，$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$，したがって

$$\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i, B)}{P(B)} = P(A_i|B)$$

が成り立つ．これが以下の**ベイズの定理**である.

定理 8 (ベイズの定理（Bayes' theorem))　たがいに背反な事象 $A_1, A_2, \cdots$, A_n が全事象 Ω を分割しているとする．このとき，事象 $B \in \mathcal{A}$ について，$P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$ が成り立つ.

例 1 昔, ある村にうそつき少年がいた．少年はいつも「オオカミが来た！！」と大声で叫んでいたが, いままで本当だったことがない.「オオカミが来た」という事象を A, 少年が「オオカミが来た！！」と叫ぶ事象を B とし, $P(B \mid A) = 1.0$, $P(B \mid A^c) = 0.5$，$P(A) = 0.005$ とする．少年が「オオカミが来た！！」と叫んだとき実際にオオカミが来ている確率を求めてみよう.

ベイズの定理より，

$$
\begin{aligned}
P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + (1 - P(A))P(B|A^c)} \\
&= \frac{0.005 \times 1.0}{0.005 \times 1.0 + 0.995 \times 0.5} \approx 0.009\,95.
\end{aligned}
$$

すなわち，うそつき少年の情報により，情報がないときに比較して，オオカミの来ている確率が約 2 倍に上昇していることがわかる．

ベイズ統計では，より広い確率の解釈として「ビリーフ」(belief) を用いることは先に述べた．ここでは考え方のみについてふれよう．意思決定問題から個人的な主観確率であるビリーフが以下のように求められる．

例えば，つぎの二つの賭けを考えよう．

1. もしオオカミが来ていれば 1 万円もらえる．
2. 赤玉 n 個，白玉 $100 - n$ 個が入っている合計 100 個の玉が入っている壺の中から一つ玉を抜き出し，それが赤玉なら 1 万円もらえる．

どちらの賭けを選ぶかといわれれば，2 番目の賭けで赤玉が 100 個ならば，誰もが迷わず 2 番目の賭けを選ぶだろうし，逆に $n = 0$ ならば，1 番目の賭けを選ぶだろう．この二つの賭けがちょうど同等になるように n を設定することができれば，$\dfrac{n}{100}$ があなたの「オオカミが来る」ビリーフになる．このように，ベイズ統計における確率の解釈「ビリーフ」は頻度主義の確率で扱える対象を拡張でき，個人的な信念やそれに基づく意思決定をも合理的に扱えるツールとなる．

ビリーフを用いてもう一度例を振り返ろう．例 1 では，もともとのオオカミが来る確率 $P(A) = 0.005$ が，(うそかどうかわからない) 少年の報告により $P(A \mid B) = 0.009\,95$ と約 2 倍にビリーフが更新されていることがわかる．すなわち，うそをつく少年の証言によって事前のビリーフが事後のビリーフに更新されたのである．このとき，ベイズ統計では，少年の証言を「**エビデンス**」(evidence) と呼び，事前のビリーフを「**事前確率**」(prior probability)，事後のビリーフを「**事後確率**」(posterior probability) と呼ぶ．

例 2 (**3 囚人問題**) つぎに有名な 3 囚人問題を紹介しよう．ある監獄にアラ

ン，バーナード，チャールズという 3 人の囚人がいて，それぞれ独房に入れられている．3 人は近く処刑される予定になっていたが，恩赦が出て 3 人のうち 1 人だけ釈放されることになったという．誰が恩赦になるかは明かされておらず，それぞれの囚人が「私は釈放されるのか?」と聞いても看守は答えない．囚人アランは一計を案じ，看守に向かって「私以外の 2 人のうち少なくとも 1 人は死刑になるはずだ．その者の名前が知りたい．私のことじゃないんだから教えてくれてもよいだろう?」と頼んだ．すると看守は「バーナードは死刑になる」と教えてくれた．それを聞いたアランは「これで釈放される確率が 1/3 から 1/2 に上がった」とひそかに喜んだ．果たしてアランが喜んだのは正しいのか?

この問題はベイズの定理を用いれば合理的に解くことができる．アランが釈放される事象を A，バーナードが釈放される事象を B，チャールズが釈放される事象を C とする．いま，事前にはだれが釈放されるかわからないので，$P(A)=P(B)=P(C)=\dfrac{1}{3}$ としてよい．また，看守はうそをつかないものとし，看守が証言した事象を E とする．ベイズの定理を用いれば，以下の事後確率を計算すればよい．

$$P(A \mid E)=\frac{P(E \mid A)P(A)}{P(E \mid A)P(A)+P(E \mid B)P(B)+P(E \mid C)P(C)}$$

$P(E \mid A)$ は，アランが釈放されるときに看守はバーナードかチャールズかどちらかの名前を言えたので，看守がバーナードが処刑されると言う確率は $\dfrac{1}{2}$ である．$P(E \mid B)$ は，看守はバーナードが釈放されるときにバーナードが処刑されるとは言わないので 0.0 である．$P(E \mid C)$ は，チャールズが釈放されるときに，アランの名前は言えないので看守はバーナードの名前しか言えず，確率は 1.0 となる．これらの値を上のベイズの定理に挿入すると

$$\begin{aligned}P(A \mid E)&=\frac{P(E \mid A)P(A)}{P(E \mid A)P(A)+P(E \mid B)P(B)+P(E \mid C)P(C)}\\&=\frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+0.0\cdot\frac{1}{3}+1.0\cdot\frac{1}{3}}=\frac{1}{3}\end{aligned}$$

結果として，アランの釈放される事後確率は事前確率と等しく，看守の証言は情報にはなっていないことがわかる．

以上のようにベイズの定理は帰納推論のための合理的なツールであることがわかる．すなわち，条件付き確率 $P(A \mid B)$ を知っていることはわれわれが因果（causality）$B \to A$ を仮定していることになる．したがって，ベイズの定理は，因果（causality）$B \to A$ を用いて A が起こったときの原因の確率 $P(B \mid A)$ を求めるツールである．そして，因果 $B \to A$ をより複雑なネットワークに拡張して，これらの推論を行うことがベイジアンネットワーク（Bayesian network）なのである．

1.5　確　率　変　数

一つの試行の結果を**標本点** $\omega \in \Omega$ と呼ぶ．この標本点 $\omega \in \Omega$ は，なんらかの測定によって観測される．この測定のことを，確率論では**確率変数**（random variable）と呼ぶ．例えば，コインを n 回投げるという試行について，表が出る回数 X は確率変数である．このとき，標本点 ω は表・裏のパターンが n 個あり得るので 2^n 通りあり，X は 0 から n までの値をとる．

数学的には，確率変数は以下のように定義される．

定義 5 確率空間 $(\Omega, \mathcal{A}, P)$ に対し，Ω から実数 $\mathcal{R}$ への関数 $X : \Omega \to \mathcal{R}$ が，任意の実数 r に対し

$$\{X \leqq r\} \in \mathcal{A}$$

を満たすならば，X を確率空間 $(\Omega, \mathcal{A}, P)$ 上の確率変数という．

また，確率変数 X が確率空間 $(\Omega, \mathcal{A}, P)$ 上に定義されると，任意の $r \in \mathcal{R}$ に対して

$$F(r) = P(\{X \leqq r\})$$

索　　　　引

—— 著 者 略 歴 ——

1989 年　神戸大学教育学部初等教育科卒業
1992 年　神戸大学大学院修士課程修了（教育学専攻）
1994 年　東京工業大学大学院博士課程早期修了（システム科学専攻）
　　　　博士（工学）
1994 年　東京工業大学助手
1996 年　千葉大学助手
2000 年　長岡技術科学大学助教授
2006 年　電気通信大学助教授
2007 年　電気通信大学准教授
　　　　現在に至る

ベイジアンネットワーク
Bayesian Network

2013 年 7 月 30 日　初版第 1 刷発行　★

検印省略	著　　者	植　野　真　臣（うえの まおみ）
	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 978–4–339–06103–1　（金）　（製本：愛千製本所）
Printed in Japan

落丁・乱丁本はお取替えいたします