Russell, S., and Norvig, P. (2021). *Artificial Intelligence: A Modern Approach (Fourth Edition)* Pearson.

第12章の要約

作成者：鈴木貴之

＊この要約は、JST/RISTEX「人と情報のエコシステム」研究開発領域プロジェクト「人と情報テクノロジーの共生のための人工知能の哲学2.0の構築」の活動として行われている研究会の資料として作成したものです。作成者の専門は人工知能ではないため、内容の誤りを含んでいる可能性があります。内容の誤りや不適切な訳語を発見した場合には、鈴木貴之 (tkykszk@g.ecc.u-tokyo.ac.jp) までお知らせいただければ幸いです。

# 第12章　不確実性を数量化する

## 12.1　不確実性下で行為する

**信念状態**の追跡によって不確実性に対処するという方法には、以下のような問題がある。

- 観察に対する可能な説明すべてを考慮する必要がある。
- 正しい contingent plan は巨大なものとなる。
- 目標達成を保証するプランがないときにエージェントが行為しなければならないこともある。

不確実性ゆえに、あるプランの成功を確実に演繹できない場合でも、何らかのプランがパフォーマンス尺度を最大化する。このような場合の**合理的決定**は、さまざまなゴールの相対的な重要性と、それぞれが達成される可能性によって決まる。

### 12.1.1　不確実性を要約する

歯科の例。歯痛と虫歯の間に確実な法則関係は見出せない。

論理学では診断を扱えない理由：

- 前件、後件の完全な記述を作成することは困難。
- 理論的に未知なことがある。
- 個々の患者について未知なことがある。

このような領域では、知識は**信念の度合い**でしかありえない。信念の度合いを扱うことができるのは**確率理論**。論理的エージェントはある文が真か偽のいずれかであると信じるが、確率的エージェントは0と1の間の信念の度合いを有する。確率理論によって不確実性を要約的に表現することが可能となる。

### 12.1.2　不確実性と合理的意思決定

エージェントはさまざまな**結果**に関する**選好**を有する。選好は**効用**で表現される。ある状態の効用はエージェントに相対的。効用の割り振り方そのものに合理／不合理はない。

**意思決定理論**＝確率理論＋効用理論

**期待効用最大化**原理：エージェントが合理的であるのは、期待効用を最大化する行為を選択するとき、そしてそのときのみ。

## 12.2　基礎的な確率の表記法

### 12.2.1　確率は何についてのものか

確率言明は可能世界についてのもの。すべての可能世界からなる集合は **標本空間**と呼ばれる。可能世界は相互に排他的であり網羅的。

完全な**確率モデル**は、すべての可能世界に確率の値 $P(\omega)$ を割り当てる。確率理論の基本公理に従えば、すべての可能世界は0から1の間の確率をもち、すべての可能世界の確率の和は1である。

通常、確率言明は可能世界の集合に関するもの。可能世界の集合は**事象** (event) と呼ばれる。

「2つのサイコロの目の和が11である確率」などは**無条件確率**または**事前確率**と呼ばれる。事前確率は、他の情報がないときのある命題に関する信念の度合いを表す。多くの場合には、われわれは**証拠**を有している。「1つ目のサイコロの目が5であるときにゾロ目が出る確率」などは**条件つき確率**または**事後確率**と呼ばれる。

意思決定の際には、エージェントは観察されたすべての証拠を条件とする必要がある。

条件 $b$ の下での $a$ の確率は、以下のように定められる。

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \tag{12.3}$$

変形すれば、

$$P(a \wedge b) = P(a|b)P(b) \tag{12.4}$$

これは**乗法定理** (the product rule) と呼ばれる。

### 12.2.2　確率言明における命題言語

本章と次章では、**因子化された表現** (factored representation) を用いて可能世界を記述する。

確率理論における変数は**確率変数** (random variables) と呼ばれる。確率変数は、可能世界からなるドメイン $\Omega$ から値域への写像。ブール的な確率変数は真と偽からなる値域をもつ。

［以下、確率変数そのものには大文字で始まる単語が、その特定の値には小文字で始まる単語が用いられている。］

確率変数に対しては**確率分布**が定義される。離散変数の確率分布はベクトルで表現できるが、連続変数の確率分布は**確率密度関数**で表現される。

複数の確率変数に対しては、**同時確率分布** (joint probability distribution) が定義される。

ある可能世界は、すべての確率変数に値を割り当てることで定義される。この定義によれば、すべての可能世界は相互に排他的かつ網羅的となる。

確率モデルは、すべての確率変数に対する同時分布、すなわち **完全な同時確率分布** (full joint probability distribution) によって完全に決定される。

### 12.2.3　確率の公理とその合理性

以上の定義から以下が導かれる。

$$P(\neg a) = 1 - P(a)$$

包除原理 (the inclusion-exclusion principle)：

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b) \tag{12.5}$$

不整合な確率的信念をもつことができないことは、以下のようにして示される。ある主体のある命題に関する信念に応じて賭けを構成することができる。de Finetti は、ある主体が確率の公理に反する信念の度合いをもつならば、その主体がつねに負ける賭けを構成できることを示した。この de Finetti の定理によれば、いかなる合理的行為者も確率公理に反する信念をもつことはできない。

## 12.3　完全な同時分布を用いた推論

観察された証拠を条件とした**クエリ**命題［＝確率を知りたい命題］の事後確率の計算方法を述べる。完全な同時分布を知識ベースとして用いる。

3つのブール変数からなるドメインを考える。ある命題の確率は、その命題が真である可能世界の確率を足すことで求められる。

| | *toothache* | | *¬toothache* | |
|---|---|---|---|---|
| | *catch* | *¬catch* | *catch* | *¬catch* |
| *cavity* | 0.108 | 0.012 | 0.072 | 0.008 |
| *¬cavity* | 0.016 | 0.064 | 0.144 | 0.576 |

**Figure 12.3** A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

ある変数に関する**周辺確率**も同様に求められる。この計算は、他の変数を式から除外する手続きであることから**周辺化**と呼ばれる。

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{Z} = \mathbf{z}) \tag{12.7}$$

（式12.7）に乗法定理を適用することで、**conditioning** と呼ばれる規則（式12.8）が得られる。

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{Z}} \mathbf{P}(\mathbf{Y}|\mathbf{z}) P(\mathbf{z}) \tag{12.8}$$

つまり、$Y$ である確率は、$Z = z$ であるという条件の下での $Y$ の確率と $Z = z$ である確率の積を、$Z$ のすべての値について加えることで求められる。

多くの場合、われわれが関心があるのは、何らかの変数の条件付き確率。条件付き確率は、（式12.3）によって条件付き確率を無条件確率に変換し、完全な同時分布を用いてその値を評価することで求められる。

一般的な推論の手続きは以下のようなものとなる。クエリが1つの変数 $X$ だけを含むとき、$\mathbf{E}$ を証拠変数のリスト、$\mathbf{e}$ を観察された値のリスト、$\mathbf{Y}$ を観察されていない変数とすると、クエリは $\mathbf{P}(X|\mathbf{e})$ であり、以下のように評価される（ $\alpha$ は標準化のための定数）。

$$\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

ドメインが $n$ 個のブール変数を含むとき、入力テーブルのサイズは $O(2^n)$ となり、処理にも同じサイズの時間がかかる。現実の状況では $n = 100$ を超えることがあるため、計算は事実上不可能となる。同時分布表を得るためには膨大な事例が必要となることも問題。

したがって、完全な同時分布の表を作成することは、推論システムを作る方法としては現実的ではない。この方法は、より実効的なアプローチのための理論的基礎と見なすべき。以下では、第13章で導入される現実的なシステムで用いられる基本的なアイデアを紹介する。

## 12.4　独立性

歯科の事例において、同時分布に第4の確率変数を導入することを考える。（式12.4）より、$P(a,b,c,d) = P(d|a,b,c)P(a,b,c)$ である。ここで、歯の問題と天気は無関係なので、$P(d|a,b,c) = P(d)$ と考えることができる。これらから、$P(a,b,c,d) = P(d)P(a,b,c)$ が導かれる。このような性質を**独立性**（絶対的独立性）と呼ぶ。独立な変数間では以下の関係が成立する。

$$P(a|b) = P(a) \quad or \quad P(b|a) = P(b) \quad or \quad P(a \land b) = P(a)P(b) \tag{12.11}$$

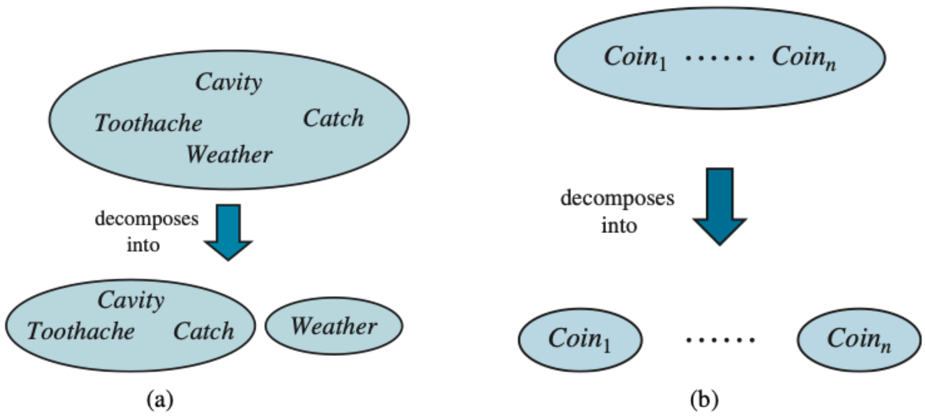


**Figure 12.4** Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.

通常、独立性はあるドメインに関する知識に基づいて主張される。独立性を想定できれば、完全な同時分布を特定するために必要な情報は大幅に減る。完全な同時分布は、相互に独立な変数の集合の同時分布に因子化されるから。

現実には、独立性によって変数を因子化できることはまれ。また、独立な部分集合自体も巨大なものとなる。

## 12.5　ベイズの規則とその使用

乗法定理は、以下のようにも表現できる。

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} \tag{12.12}$$

この式は**ベイズの規則**（ベイズの定理）と呼ばれる。現代の人工知能による確率的推論の基礎となるのは、この規則。この式を何らかの背景的証拠 $\mathbf{e}$ によって条件付ければ、（式12.13）となる。

$$\mathbf{P}(Y|X,\mathbf{e}) = \frac{\mathbf{P}(X|Y,\mathbf{e})\mathbf{P}(Y|\mathbf{e})}{\mathbf{P}(X|\mathbf{e})} \tag{12.13}$$

### 12.5.1　ベイズの規則の適用：単純な事例

ベイズの規則が有用であるのは、$P(a|b), P(b), P(a)$ が推測可能で、そこから $P(b|a)$ を計算する必要がある場面がしばしばあるから。たとえば、観察された結果から原因を推測するときには、ベイズの規則から以下が成り立つ。

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

$P(\mathit{effect}|\mathit{cause})$ は**因果**方向の関係であるのに対して、$P(\mathit{cause}|\mathit{effect})$ は**診断**方向の関係。前者については、われわれはしばしば知識を有する。

12.3節では、事前確率 $P(b)$ が分からないときには、事後確率を知りたい確率変数のすべての値について事後確率を求め、それを標準化するという手法を紹介した。ベイズの規則を用いる際にも同様の手法を利用できる。

なぜある方向の条件付き確率の知識だけをもつことがあるのか？診断的な確率は統計的に得られるが、状況に応じて変化する。これに対して、因果的な確率はそうでない。因果的な知識はより頑健。

### 12.5.2　ベイズの規則の使用：証拠を結合する

複数の証拠を用いる場合にはどうするか？ $n$ 個の証拠があるときには $2^n$ 通りの条件付き確率を知る必要が生じるため、現実的ではない。

ドメインにさらなる仮定を設けることで問題を単純化できる。確率変数が独立であればよいが、そうであるとはかぎらない。しかし、虫歯がある（ない）という条件の下では、歯痛と触診結果は独立となる。どちらも虫歯によって引き起こされるが、相互に因果的影響をもたないから。つまり以下のような関係が成り立つ。

$$\mathbf{P}(\mathit{toothache} \wedge \mathit{catch}|\mathit{Cavity}) = \mathbf{P}(\mathit{toothache}|\mathit{Cavity})\mathbf{P}(\mathit{catch}|\mathit{Cavity})$$

この式は、虫歯によって条件付けられた歯痛と触診結果の**条件付き独立性**を表現している。これを用いると、（式12.18）が得られる。

$$\mathbf{P}(\mathit{Cavity}|\mathit{toothache} \wedge \mathit{catch}) = \alpha\mathbf{P}(\mathit{toothache}|\mathit{Cavity})\mathbf{P}(\mathit{catch}|\mathit{Cavity})\mathbf{P}(\mathit{Cavity}) \quad (12.18)$$

一般的に、2つの確率変数 $X$ と $Y$ の第3の変数 $Z$ を条件とした条件付き独立性は以下のように表現できる。

$$\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z)\mathbf{P}(Y|Z)$$

条件付き独立言明についても、12.4節で論じたのと同様の因子化が可能。この操作の結果、必要とされる表は小規模なものとなる。

$n$ 個の症状が虫歯を条件として条件付き独立な場合には、必要となる情報は $O(2^n)$ ではなく $O(n)$ のオーダーで増加する。つまり、条件付き独立言明によって、確率的システムは規模を拡大できるようになる (scalable) 。また、条件付き独立言明は絶対的独立言明よりも利用可能なことが多い。巨大な確率的ドメインを、条件付き独立性によって弱く結合された部分集合に分解する手法は、人工知能研究の近年の動向におけるもっとも重要な発展の一つ。

## 12.6　ナイーブベイズモデル

歯科の事例は、単一の原因が条件付き独立な複数の結果に影響するパターンの一例。このような場合の完全な同時分布は、以下のように表現できる。

$$\mathbf{P}(\mathit{Cause}, \mathit{Effect}_1, \ldots, \mathit{Effect}_n) = \mathbf{P}(\mathit{Cause})\prod_i \mathbf{P}(\mathit{Effect}_i|\mathit{Cause}) \quad (12.20)$$

このような分布で表現されるモデルは**ナイーブベイズ** モデルと呼ばれる。原因である変数を条件としたときに結果となる変数が厳密には独立でない場合にしばしば用いられるので、「ナイーブ」と呼ばれる。そのような場合でも、ナイーブベイズシステムはしばしばうまく働く。

（式12.20）を用いて、観察された諸結果を条件とした原因の確率を得ることができる。観察された諸結果を $\mathbf{E} = \mathbf{e}$ と呼び、観察されていない諸結果を $\mathbf{Y}$ と呼ぶ。（式12.9）を用いると、

$$\mathbf{P}(\mathit{Cause}|\mathbf{e}) = \alpha \sum_y \mathbf{P}(\mathit{Cause}, \mathbf{e}, \mathbf{y})$$

（式12.20）を適用すると次の（式12.21）が得られる。

$$\mathbf{P}(Cause|\mathbf{e}) = \alpha \mathbf{P}(Cause) \prod_j \mathbf{P}(e_j|Cause) \tag{12.21}$$

言葉で表すとつぎのようになる。可能なすべての原因に対して、原因の事前確率と、原因によって条件付けられた観察された結果の条件付き確率の積を掛けあわせ、その結果を標準化する。この計算に要する時間は、観察された結果の数に線形に比例し、観察されていない結果の数には影響されない。このように、値が観察されていない変数が計算から消え去ることは、確率的推論では一般的。

### 12.6.1　ナイーブベイズによるテキストの分類

新聞記事をカテゴリーに分類するという事例を考える。原因はカテゴリー、結果はあるキーワードの有無。ナイーブベイズモデルは、$\mathbf{P}(Category)$ と $\mathbf{P}(HasWord_i|Category)$ からなる。$P(Category = c)$ や $\mathbf{P}(HasWord_i|Category)$ は過去の記事から求められる。

ナイーブベイズモデルは、文書において語が相互に独立に出現すると仮定している。これは事実とは異なる。このような場合、モデルから予測される事後確率は1または0に近くなりすぎることになるが、カテゴリーの相対的順位には影響はない。

ナイーブベイズモデルは言語の特定、文書の呼び出し、スパムのフィルタリングなどに用いられている。事後確率の値そのものが重要な医療における診断などでは、次章で紹介する、より洗練されたモデルが用いられる。

## 12.7　The Wumpus World再訪

［省略］

最終更新：2022年1月22日