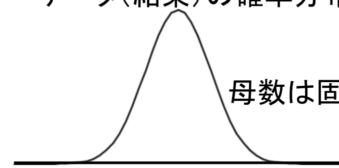
ベイズ統計の概要

2021年　ver. 1.0

倉谷　隆博

12－1. ベイズの定理（1/2）

（表1）頻度主義とベイズ主義の比較

	頻度主義	ベイズ主義
確率	原因 → 結果	結果 → 原因
確率分布	データ（結果）の確率分布 	母数（原因）の確率分布 

（表1）に従来の頻度主義とベイズ主義を比較する。
頻度主義は、ランダムな事象が発生する頻度をもって確率とする考え方であり、ある原因のもとである結果が得られる確率を考える。つまり、原因から結果の方向、因果関係で確率を考える。
一方、ベイズ主義は、ある結果が得られたときその結果をもたらした原因の確率を考える。つまり、結果から原因へ時間をさかのぼって原因の確率を考える。
この確率の考え方は確率分布へも拡大される。原因を確率分布を規定する母数（平均、分散など）に、結果をこの確率分布に従って発生したデータに置き換える。
頻度主義では、母数（原因）が規定する確率分布に従って発生するデータ（結果）を分析してデータの特性を把握する。一方、ベイズ主義では、母数自身が確率分布を持つと考え、発生したデータ（結果）をもとに母数の確率分布（原因）を調べる。

ベイズ統計の土台になるベイズの定理は、下記の確率の「乗法定理」から導かれる。

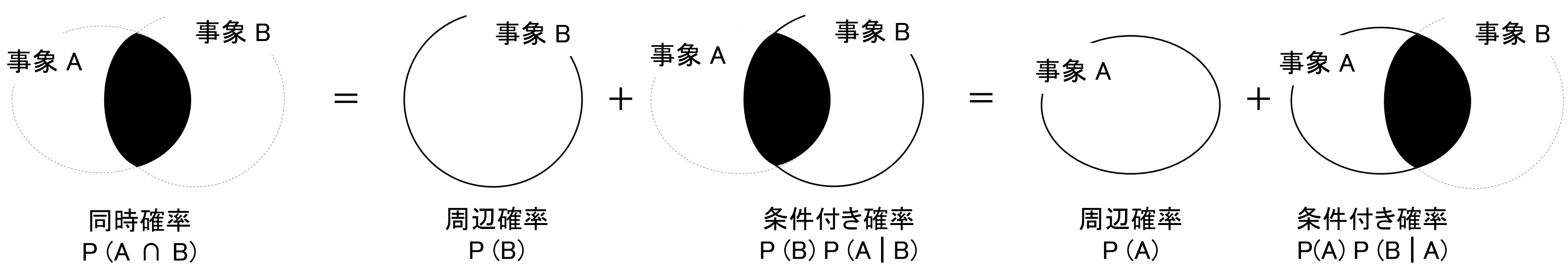
1）乗法定理

$$P(A \cap B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A) \quad \cdots \quad (式1)$$

ここで、

$P(A \cap B)$ ： $A \cap B$ は事象 A と B が同時に起こるという事象であり、その確率は同時確率という
$P(B)$ ： 事象 B が起こる確率、周辺確率という
$P(A \mid B)$ ： $A \mid B$ は事象 B が起こったという条件のもとで 事象 A が起こるという事象であり、その確率は条件付き確率という

ベン図を使うと、乗法定理（式1）は（図1）のように表すことができる。



（図1）乗法定理

12－1. ベイズの定理（2/2）

2）ベイズの定理

乗法定理（式1）からベイズの定理（式2）が求まる。

$P(A \mid B) = P(B \mid A) \cdot P(A) / P(B)$ ・・・（式2）

ここで、事象 A を「原因、仮説」、事象 B を「結果、データ」と解釈することが、ベイズ統計学の誕生につながった。つまり（表2）に示す通り、事後確率と称する条件付き確率 $P(A \mid B)$ は、事前確率と称する $P(A)$ と、尤度と称する条件付き確率 $P(B \mid A)$ 、および確率 $P(B)$ から求まる（尤度とは、観測している事象 = 結果が起こりうる確率のこと）。なお、（式2）は原因の確率を計算する式であるが、分母の $P(B)$ は原因の有無を問わない確率であり、この式では定数であるとみなすことができる。

（表2）ベイズの定理の説明

確率	名称	説明
P（A｜B）	事後確率	結果B が得られたときに、その原因が A である確率
P（B｜A）	尤度（ゆうど）	原因Aのもとで結果Bが得られる確率
P（A）	事前確率	結果Bが得られる前の原因Aが発生する確率
P(B)	定数扱い	原因Aの有無にかかわらず結果Bが発生する確率

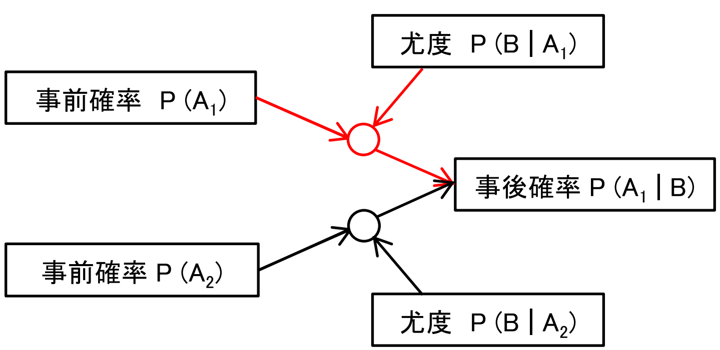
ベイズの定理は、原因が複数ある場合には以下のように拡張される。
下式は、原因が2つの場合で、かつ事象 A_1、A_2（原因）が相互干渉しない、つまり独立している場合に成り立つ。

$P(A_1 \mid B) = P(B \mid A_1) \cdot P(A_1) / P(B)$ ・・・（式3－1）

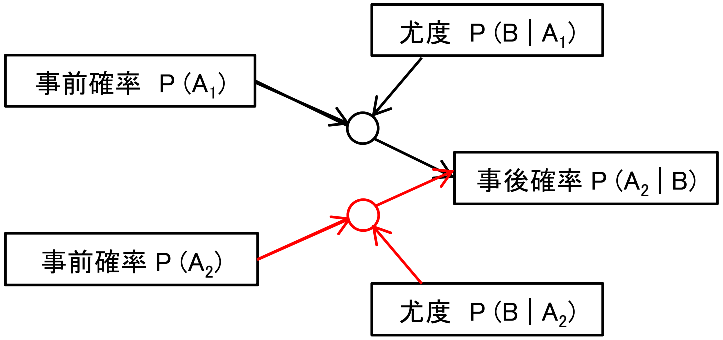
$P(A_2 \mid B) = P(B \mid A_2) \cdot P(A_2) / P(B)$ ・・・（式3－2）

なお、$P(B)$ は両式に共通で、$P(B) = P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2)$ ＝ 定数　である。

一般には、 $P(A_i \mid B_j) = P(B_j \mid A_i)\ P(A_i) / \Sigma_k P(B_j \mid A_k) P(A_k)$ ・・・（式4） である。



（図2）結果 B から、原因 A_1 の事後確率を求める場合

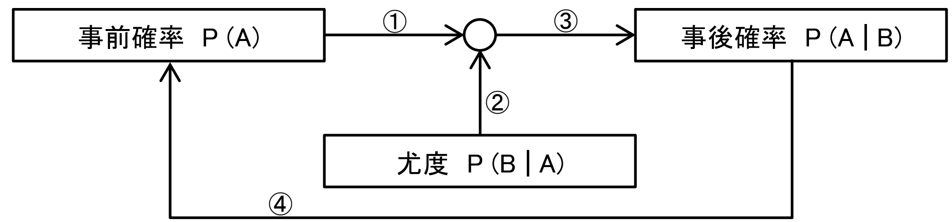


（図3）結果 B から、原因 A_2 の事後確率を求める場合

12−2. ベイズ更新(オオカミ少年の話)

ベイズ統計を特徴づける考え方の一つに「ベイズ更新」がある。これは(図4)に示すように、① 事前確率をもとに、② これに入手した情報(尤度)を加味して、③ 事後確率を求める。次いで、次の新しい情報(尤度)を入手したら、④ 先に求めた事後確率を事前確率に置き換えて、② 新たな情報(尤度)を加味して、③ 事後確率を計算し直す。こうしたサイクルがベイズ更新であり、新たな情報が入手されるたびにベイズ更新を繰り返し行って真の事後確率の値に近づいていく。
なお、事前確率を仮の値に設定して更新を始めることができるのもこの方法の強みである。仮の設定値は主観的な値で良く、理由不十分の原則(例えば、2者択一の確率は情報が不十分のときは、取り敢えず等確率であるとする考え方)を採用する。

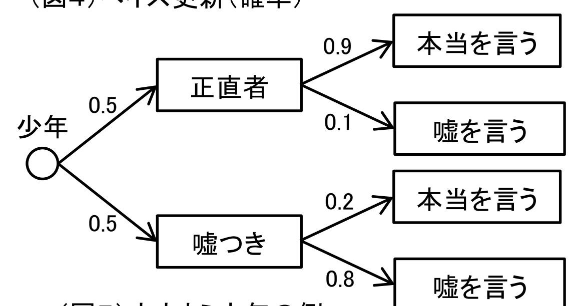
ベイズ更新の考え方は、一般の統計学の考え方とは全く異なる。一般の統計学(頻度主義)では、データをもとに母集団の特性を計算し、新たなデータが発生したときには計算し直す。一方、ベイズ統計では、ベイズ更新の考え方を使って、新たなデータだけを加味する方法をとる。これは、過去に学んだ知識の上に新たな情報を追加していく「学習」プロセスそのものであり、人工知能(AI)の分野にベイズ統計の考え方が利用されている所以でもある。



P(A | B) = P(B | A)・P(A) / P(B)
事後確率 = 尤度・事前確率 / 定数

(図4)ベイズ更新(確率)

■オオカミ少年の話　(例題)
オオカミ少年はどの程度の嘘つきなのかを、嘘を言った事実から判断する(図5)。
嘘つきであれば 0.8 の確率で嘘を言い、0.2 の確率で本当を言う、
正直者であれば 0.9 の確率で本当を言い、0.1の確率で嘘を言うとする。
また、最初は嘘つきの程度が分からないので、嘘つきの確率はとりあえず0.5であるとする(この仮定を許容することを「理由不十分の原則」という)。



(図5)オオカミ少年の例

ここで、原因 = 嘘つき、結果=嘘を言う、と置いて、
尤度 = P(嘘を言う | 嘘つき) = 0.8 、事前確率 = P(嘘つき) = 0.5 、
定数 = P(嘘を言う) = P(正直者であろうが嘘つきであろうが嘘を言う) = 0.5 × 0.1 + 0.5 × 0.8 = 0.45 であるので、
P(嘘つき | 嘘を言う) 事後確率 = 尤度・事前確率 / 定数 = 0.8 × 0.5 / 0.45 = 0.889
この結果から少年が嘘つきである確率は当初 0.5 としたが、0.889 に更新された。

ここで少年がさらに嘘を重ねたら、P(嘘つき | 嘘を言う) 事後確率 = 0.8 × 0.889 / ((1 − 0.889) × 0.1 + 0.889 × 0.8) = 0.985 になり、嘘つきの確率が上がる。

次に、ここで少年がようやく本当を言ったとする。しかし、
P(嘘つき | 本当を言う) 事後確率 = 0.2 × 0.985 / ((1 − 0.985) × 0.9 + 0.985 × 0.2) = 0.936 になり、
少年が嘘つきである確率は多少下がるだけで、嘘を繰り返していたので、本当を言っても高い確率で嘘つきだと判断されてしまう。

１２－３. ナイーブベイズ判別分析（naive Bayes filter）

ナイーブベイズフィルターはベイズ統計の事後確率を用いた判別分析であり、線形判別分析ではどちらのカテゴリーに属するのか判別が難しいようなデータについて、事後確率を使ってデータが所属する確率が高いカテゴリーを判別する方法である。

ナイーブベイズ判別（単純ベイズ判別）では、データの間には相関がないとした上で、ベイズ更新の手法を使う。

カテゴリー（原因として扱う）が A1 と A2 の２つあり、データ（結果として扱う）が B1, B2,・・・ であるとき、

最初の結果 B_1 が得られる原因が A_1 である確率（事後確率、確信度）は、（式３－１）から

$$P(A_1 \mid B1) = \text{定数} \cdot P(B_1 \mid A_1) \cdot P(A_1) \quad \cdots \quad (\text{式}5-1)$$

引き続き、次の結果 B_2 が得られる原因が A_1 である確率は、（式５－１）を使って、つまりベイズ更新を使って、

$$P(A_1 \mid B_2, B_1) = \text{定数} \cdot P(B_2 \mid A_1) \cdot P(A_1 \mid B_1) = \text{定数} \cdot P(B_2 \mid A_1) \cdot \{\text{定数} \cdot P(B_1 \mid A_1) \cdot P(A_1)\}$$
$$= \text{定数} \cdot P(B_2 \mid A_1) \cdot P(B_1 \mid A_1) \cdot P(A_1) \quad \cdots \quad (\text{式}5-2)$$

この式から、事前確率 $P(A_1)$ に、新しい結果が得られる度に、その結果（尤度、出現確率）として、最初は $P(B_1 \mid A_1)$ を、次いで $P(B_2 \mid A_1)$を加味していけば（掛けていけば）、原因がA_1である確率が求まる。原因 A2 の確率に関しても同様である。

■スパムメールの判別　（例題）

ナイーブベイズ判別分析は、スパムメールの検出への応用でよく知られている。メールから抽出する単語（データ、結果）を使ってスパムメールであるか、正常メール（ハムメールとも言う）であるか（カテゴリー、原因）を判別する。

スパムメールを A_1, ハムメールを A_2 として、（式５－２）の尤度 $P(B_1 \mid A_1)$, $P(B_2 \mid A_1)$, ・・・ は過去のデータをもとに、例えば（表３）のように事前に求めておく。スパムメールの確率 $P(A_1)$ とハムメールの確率 $P(A_2)$ も、例えば（表４）のように過去のデータをもとに設定しておく。

スパムメールの事後確率（確信度） $P(A_1 \mid B_1)$, $P(A_1 \mid B_2, B_1)$は、（式５－１）、（式５－２）を使って求める。メールから単語を抽出するごとに計算を更新する。同様に、ハムメールの事後確率（確信度） $P(A_2 \mid B_1)$, $P(A_2 \mid B_2, B_1)$ も求めることができる。そして、それぞれの事後確率を比較して、事後確率の高いほうのカテゴリーを選択する（スパムメールか否かを判別する）。

（表３）スパムメール判別のための尤度

抽出した単語（結果）	スパムメール（原因　A_1）での単語の出現確率 尤度 $P(B \mid A_1)$	ハムメール（原因　A_2）での単語の出現確率 尤度 $P(B \mid A_2)$
B_1 ：「評判」	0.14	0.73
B_2 ：「招待」	0.43	0.52
B_3 ：「連絡先」	0.50	0.25
B_4 ：「出会い」	0.88	0.01

（表４）スパムメール判別のための事前確率

	スパムメールである確率 $P(A_1)$	ハムメールである確率 $P(A_2)$
事前確率	0.3	0.7

それでは、メールから単語「評判」、「招待」、「連絡先」を抽出したとして、このメールがスパムメールである確信度を計算する。

$P(A_1 \mid B_3, B_2, B_1) = (0.14 \times 0.43 \times 0.50) \times 0.3 = 0.0090$ ・・・ ①

一方、このメールがハムメールである確信度は、

$P(A_2 \mid B_3, B_2, B_1) = (0.73 \times 0.52 \times 0.25) \times 0.7 = 0.0664$ ・・・ ②

この結果（① ＜ ②）から、ハムメールである確信度の方が高いので、このメールはハムメールであると判別できる。

次いでこの後、単語「出会い」を抽出すると、事後確率は更新されて、

$P(A_1 \mid B_4, B_3, B_2, B_1) = 0.0090 \times 0.88 = 0.0080$ ・・・ ③

$P(A_2 \mid B_4, B_3, B_2, B_1) = 0.0664 \times 0.01 = 0.0007$ ・・・ ④

よって、新たな単語「出会い」を抽出したことで、ハムメールと判別していた結果が ③ ＞ ④ であるので、スパムメールという判別に変わった。

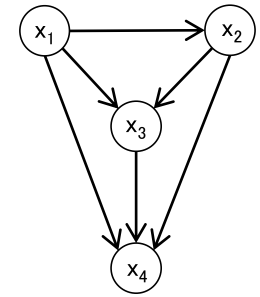
12－4. ベイジアン ネットワーク

ベイジアンネットワークの活用分野は幅広い。例えば、マーケティングではウェブ閲覧履歴を分析してお薦め商品を提示したり、製造の現場では不具合発生とその原因との関係をモデル化することができる。ベイジアンネットワークは複数の「原因」と「結果」という因果関係をネットワーク（グラフ）を使った図と確率で可視化する。確率変数間の因果関係を視覚的に表現する方法である。ただし、因果関係の向きは人の判断に委ねる。

乗法の定理（式1）の事象 A, B をそれぞれ原因 x_1 と結果 x_2 という確率変数に置き換えて（式6）を得る。x_1 の値を知った上で（条件付き確率）、x_2 が起こる確率、つまり原因と結果と関係（同時確率、因果関係）を求める式になっている。この式は、因果関係の連鎖である x_1（原因）→ x_2（結果 = 原因）→ x_3（結果 = 原因）→ x_4（結果）を表す（式7）に拡張できる。この関係は（図6）のように表すことができる。

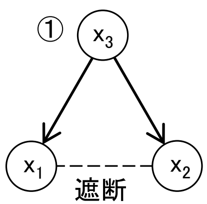
$$p(x_1, x_2) = p(x_2 | x_1)\ p(x_1) \quad \cdots \quad (式6)$$

$$p(x_1, x_2, x_3, x_4) = p(x_4 | x_3, x_2, x_1)\ p(x_3 | x_2, x_1)\ p(x_2 | x_1)\ p(x_1) \quad \cdots \quad (式7)$$



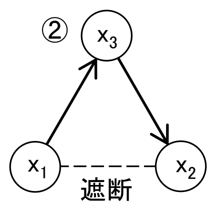
（図6）因果関係グラフ

ただ実際は、（図6）のように全ての変数間に因果関係は存在するわけではない。この因果関係の強さと有無を調べるにあたり、ある変数が観測されたという条件の下で、それに関連する2つの変数が互いに独立（条件付き独立という）であるか否かを調べ、条件付き独立が成立すれば依存関係が存在しない、成立しなければ相関関係があると判断する方法がある（図7）。

①


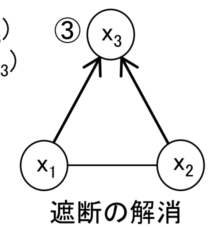
$p(x_1, x_2, x_3) = p(x_1, x_2, x_3) / p(x_3)$
$= p(x_1 | x_3)\ p(x_2 | x_3)\ p(x_3) / p(x_3)$
$= p(x_1 | x_3)\ p(x_2 | x_3)$

X_3 が観測されると、x_1 と x_2 は独立になる。つまり条件付き独立が成立し、x_1 と x_2 は関係が遮断されたという。

②


$p(x_1, x_2, x_3) = p(x_1, x_2, x_3) / p(x_3)$
$= p(x_2 | x_3)\ p(x_3 | x_1)\ p(x_1) / p(x_3)$
$= p(x_2 | x_3)\ p(x_1, x_3)\ p(x_3)$
$= p(x_2 | x_3)\ p(x_1 | x_3)$

①と同じく、x_1 と x_2 の関係は遮断され、関係がなくなる。

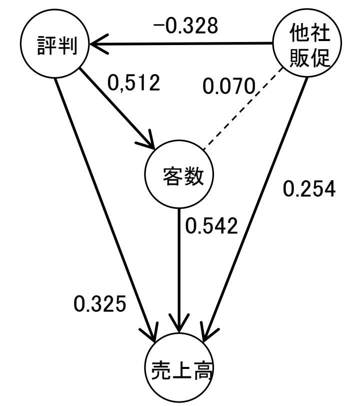
③


$p(x_1, x_2, x_3) = p(x_1, x_2, x_3) / p(x_3)$
$= p(x_3 | x_1, x_2)\ p(x_1)\ p(x_2) / p(x_3)$
$\neq p(x_1 | x_3)\ p(x_2 | x_3)$

①、②と異なり、x_3 が観測されると、独立であった x_1 と x_2 との間に関係が生まれる（遮断関係が解消される）。

（図7）条件付き独立の判断

しかし、実際には取り扱う変数の数は多く、変数間の関係は指数的に増える。
そこで、2つの変数間の偏相関係数を網羅的に調べ、偏相関係数が 0 のときには条件付き独立が成立するという関係を利用し、偏相関係数がほぼ 0 に近い値をとるとき、この2つの変数の間には実質的な関係は存在しないと判断するという方法が使われる。
（付録1）に計算例の詳細を示すが、（図8）は正規分布に従う4変数のベイジアンネットワークの事例である。変数の間の因果関係の強さを示す偏相関係数の値は図中に示す。図では客数と他社販促の間の偏相関係数の値は小さいので、この2つの変数は条件付き独立を満たす、因果関係はないと判断し、破線表示にした。一方、偏相関係数の大きな値のところを辿ると、評判が客数に影響を与え、次いで売上高に影響を与えているという関係が見える。



（図8）因果関係グラフの簡略化

12－5. ベイズの基本公式

（表5）の関係を使って、確率の関係「ベイズの定理」（式2）をもとに、確率分布の関係「ベイズの基本公式」（式8）を得る。

$w(\theta|D) = \{f(D|\theta) \cdot w(\theta)\} / P(D)$ ・・・ （式8）

$P(D) = \int f(D|\theta) \cdot w(\theta)\, d\theta$ （なお、∫は積分記号）

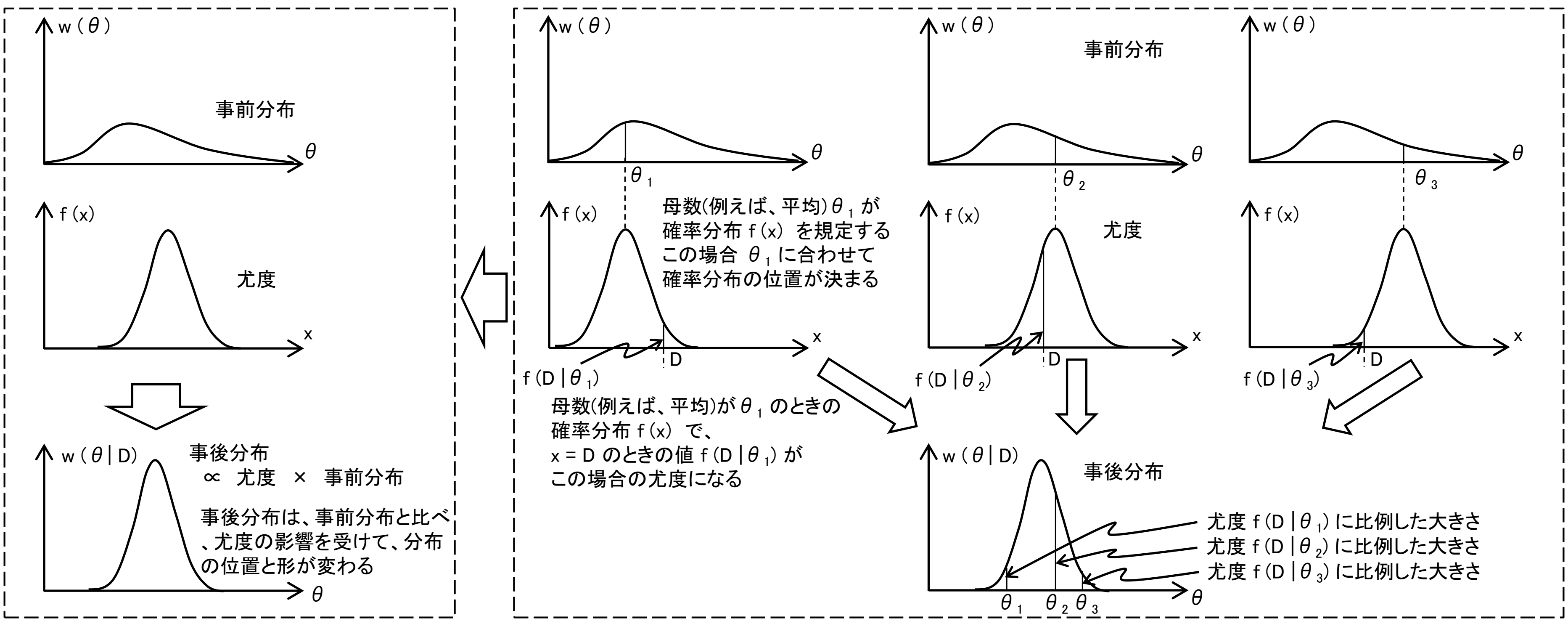
ここで、P (D) は定数なので、比例記号 ∝ を使えば、

$w(\theta|D) \propto f(D|\theta) \cdot w(\theta)$ ・・・ （式9）

（式9）は、母数 θ の確率分布（事前分布） w (θ) に、その母数 θ が規定する確率分布 f (x) のもとで発生したデータ x = D の確率である尤度 f (D |θ) を反映すると、母数 θ の新たな確率分布（事後分布） w (θ| D) を求めることができる、ことを示している。（図9）に（式9）の模式図を示す。

（表5）「ベイズの定理」と」「ベイズの公式」の」対比

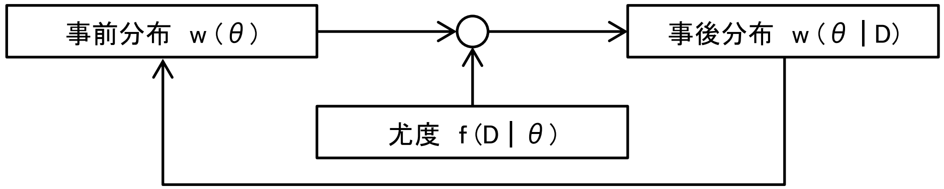
確率 「ベイズの定理」		確率分布 「ベイズの基本公式」	
記号	名前	記号	名前
A	原因	θ	母数（平均、分散などのパラメータ）
B	結果	D	データ
P (A)	事前確率	w (θ)	事前分布（確率分布）
P (A | B)	尤度	f (D | θ)	尤度（確率分布 f (x) の値）
P (B | A)	事後確率	w (θ | D)	事後分布（確率分布）
P (B)	周辺確率	P (D)	周辺尤度（事後分布の規格化定数）



（図9）ベイズの基本公式の模式図　　右図は母数 θ の値3つに関して模式図化したものであり、それらを統合したものが左図になる

12－6. 事後分布のベイズ更新（1/2）

確率分布のベイズ更新（図10）は確率のベイズ更新（図4）の考え方と同じである。
事前分布に最新の情報だけを反映して更新すれば事後分布を得ることができる。つまり、更新にあたっては過去の情報に遡る必要がない。また、データ同志が独立であれば更新順序は問わない（逐次合理性という）。



$w(\theta \mid D) = f(D \mid \theta) \cdot w(\theta) / P(D)$

事後分布 ＝ 尤度・事前分布 / 定数

（図10）ベイズ更新（確率分布）

■勝率の確率分布の例

野球の試合でA チームと B チームは3試合対戦し、結果はA チームの2勝1敗であった。A チーム の勝率の確率分布 θ を求める。

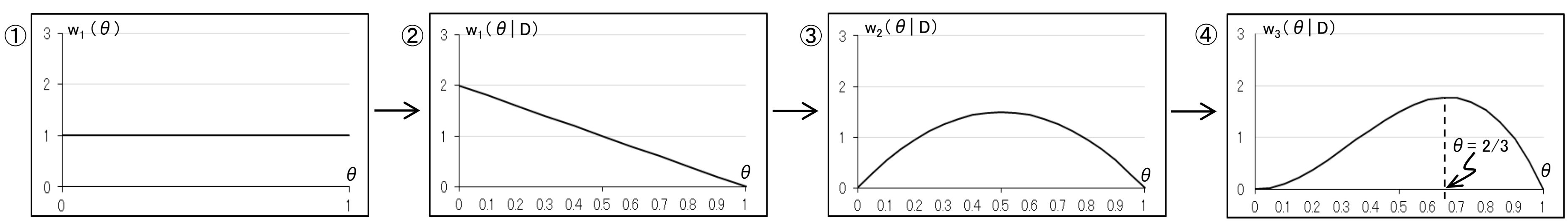
・ケース（1）　A チーム： 第1回戦＝敗戦、第2回戦＝勝利、第3回戦＝勝利　の場合

第1回戦のときには、A チームの勝率 θ に関する情報がないので、勝率 θ の事前分布を一様分布に設定する（理由不十分の原則を適用する）。（表6－1）は3戦の結果をその都度、事前分布へ反映して求めた勝率 θ の事後分布である。（図11－1）にベイズ更新の様子を、事前分布 ① から始まり、事後分布 ②、③、④ と更新されていく様子を示す。なお、第3回戦後の事後分布の最大確率を与える θ を第3回戦終了後の勝率の推定値とする（MAP推定法という）。この推定値は（図11－1）④の通り $\theta = 2/3$ であり、3回戦で2回勝利したので当然の値であるが、ピンポイントで推定（点推定）しているのではなく、確率分布のなかで他の値である確率も示しながら最も確率の高い値を選ぶのが MAP 推定法である。

（注）MAP　： Maximum A posteriori Probability estimation methods

（表6－1）　事後分布を求める計算（ケース1の場合）

ケース（1）	事前分布 （ここでは、ベータ分布）	尤度 （ここでは、ベルヌーイ分布）	周辺尤度（積分）	事後分布 （ここでは、ベータ分布）
第1回戦 A チーム 敗戦	① $w_1(\theta) = 1$ （一様分布）	$f(D \mid \theta)$ $= f(\text{敗戦} \mid \theta) = f(0) = 1 - \theta$	$\int f(D \mid \theta) \cdot w_1(\theta)\, d\theta$ $= \int (1 - \theta) \cdot 1\, d\theta = 1/2$	② $w_1(\theta \mid D) = w_1(\theta \mid \text{敗戦})$ $= (1 - \theta) \cdot 1 / (1/2) = 2(1 - \theta)$
第2回戦 A チーム 勝利	$w_2(\theta) = 2(1 - \theta)$	$f(D \mid \theta)$ $= f(\text{勝利} \mid \theta) = f(1) = \theta$	$\int f(D \mid \theta) \cdot w_2(\theta)\, d\theta$ $= \int \theta \cdot 2(1 - \theta)\, d\theta = 1/3$	③ $w_2(\theta \mid D) = w_2(\theta \mid \text{勝利})$ $= \theta \cdot 2(1 - \theta) / (1/3) = 6(\theta - \theta^2)$
第3回戦 A チーム 勝利	$w_3(\theta) = 6(\theta - \theta^2)$	$f(D \mid \theta)$ $= f(\text{勝利} \mid \theta) = f(1) = \theta$	$\int f(D \mid \theta) \cdot w_3(\theta)\, d\theta$ $= \int \theta \cdot 6(\theta - \theta^2)\, d\theta = 1/2$	④ $w_3(\theta \mid D) = w_3(\theta \mid \text{敗戦})$ $= \theta \cdot 6(\theta - \theta^2) / (1/2) = 12(1 - \theta)\theta^2$



（図11－1）　ケース（1）のベイズ更新の様子

12－6. 事後分布のベイズ更新 （2/2）

■勝率の確率分布の例 （続き）

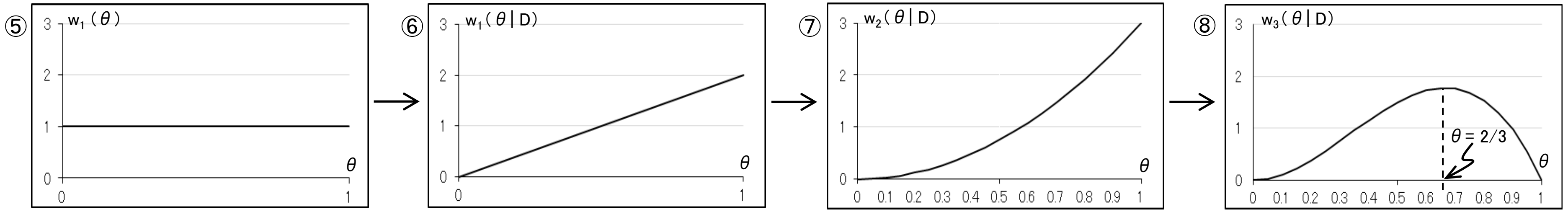
・ケース（2）　 A チーム： 第1回戦＝勝利、第2回戦＝勝利、第3回戦＝敗戦　の場合

ケース（1）のときと同様に事後分布のベイズ更新を行なった結果が（表6－2）である。（図11－2）にベイズ更新の様子を示す。

ケース（1）とケース（2）は勝敗の順序は異なるが勝敗数は2勝1敗と同じである。事後分布 ④ と事後分布 ⑧ は同じになる（逐次合理性）。

（表6－2） 事後分布を求める計算（ケース1の場合）

ケース（1）	事前分布	尤度	周辺尤度（積分）	事後分布
第1回戦 A チーム 勝利	⑤ $w_1(\theta)=1$	$f(D \mid \theta)$ $=f($勝利$\mid \theta)=f(1)=\theta$	$\int f(D \mid \theta)\cdot w_1(\theta)\,d\theta$ $=\int \theta \cdot 1\,d\theta = 1/2$	⑥ $w_1(\theta \mid D)=w_1(\theta \mid$ 勝戦$)$ $=\theta \cdot 1/(1/2)=2\theta$
第2回戦 A チーム 勝利	$w_2(\theta)=2\theta$	$f(D \mid \theta)$ $=f($勝利$\mid \theta)=f(1)=\theta$	$\int f(D \mid \theta)\cdot w_2(\theta)\,d\theta$ $=\int \theta \cdot 2\theta\,d\theta = 2/3$	⑦ $w_2(\theta \mid D)=w_2(\theta \mid$ 勝利$)$ $=\theta \cdot 2\theta/(2/3)=3\theta^2$
第3回戦 A チーム 敗戦	$w_3(\theta)=3\theta^2$	$f(D \mid \theta)$ $=f($敗戦$\mid \theta)=f(0)=1-\theta$	$\int f(D \mid \theta)\cdot w_3(\theta)\,d\theta$ $=\int (1-\theta)\cdot 3\theta^2\,d\theta = 1/4$	⑧ $w_3(\theta \mid D)=w_3(\theta \mid$ 敗戦$)$ $=(1-\theta)\cdot 3\theta^2/(1/4)=12(1-\theta)\theta^2$



（図11－2） ケース（2）のベイズ更新の様子

12－7. 事後分布を求める方法 （共役事前分布）

（1）解析的に解く方法

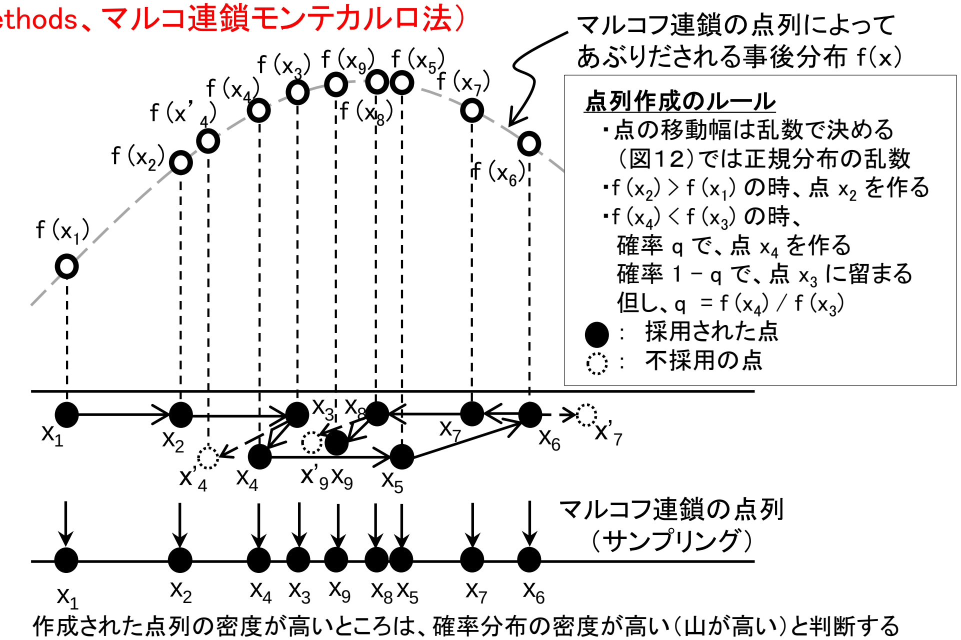
ベイズの基本公式を使って、尤度と事前分布を使って求める。

特定の尤度に対して事後分布が求まるように事前分布を決める。この事前分布のことを共役事前分布と言い、事後分布は共役事前分布と同じタイプの分布になる。例えば12－6の事例では、尤度（母数が規定する確率分布）がベルヌーイ分布（確率変数が 0 と 1 の2値だけをとる確率分布）であり、共役事前分布としてベータ分布（一様分布はベータ分布の特殊な場合）を設定すれば事後分布もベータ分布になる。他にも、尤度が二項分布（確率が 0 か 1 かの試行を反復するときの確率分布）のときの共役事前分布は、ベータ分布になるなどが知られている。しかし、この方法は共役事前分布を設定できる場合に限定される。実用的な方法ではない。

12－7. 事後分布を求める方法 （MCMC法）

2）数値的に解く方法 ： MCMC法 （Markov Chain Monte Carlo methods、マルコ連鎖モンテカルロ法）

ベイズの基本公式は、母数の数が複数であったり、確率分布が複雑であったりすると、積分の項があるため解析的に解くことはほぼ不可能になる。そこで、数値解析的に事後分布を求める方法として、MCMC 法が考案された。事後分布の値に比例した密度で分布する点列（マルコフ連鎖の点列）を作り出した上で、点列の総和を近似計算して（モンテカルロ法）事後分布の姿をあぶり出す方法である。ここで、マルコフ連鎖とは1つ前の状態だけで次の状態が決まる点列であり、モンテカルロ法とは近似計算法である。なお、確率分布に沿ったマルコフ連鎖の点列の作成方法（サンプリング方法）の一つが（図12）に示すメトロポリス法である。このサンプリングロジックによってマルコフ連鎖の点列の密度から確率分布の姿があぶりだされる。

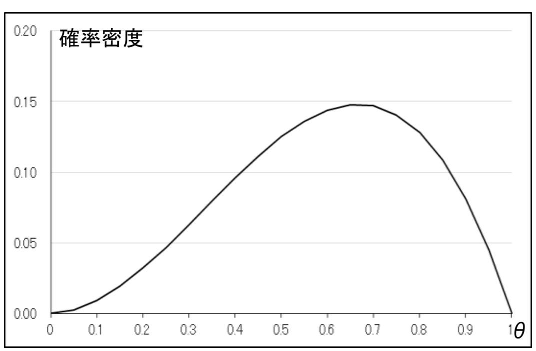


（図12）メトロポリス法（MCMC法）

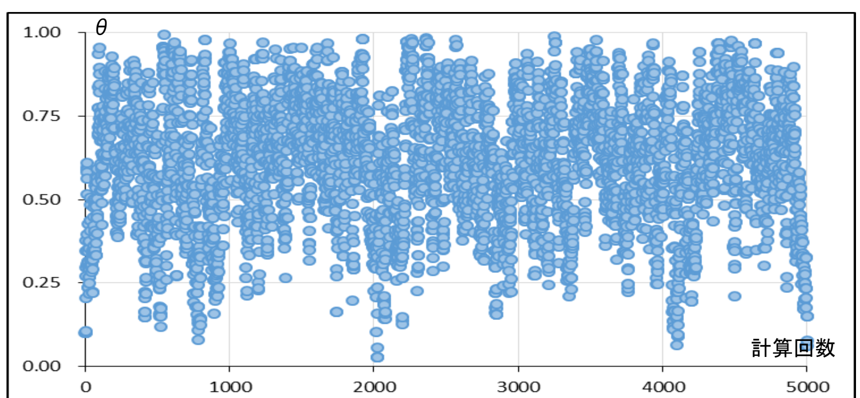
■MCMC 法の事例 （メトロポリス法） Excelで作成

あぶりだすべき事後分布を（図13）に示す。（図14）は 5,000個 のマルコフ連鎖の点列に対する事後分布の値である。（図15）は（図14）の値の大きさで整理した頻度のグラフである。ウォームアップ期間（バーンイン期間）として最初のデータ 1,000 個 は除いて（図15）を作成している。（図15）の形は（図13）の事後分布の形に近く、マルコフ連鎖の点列が事後分布をあぶりだしていることが分かる。

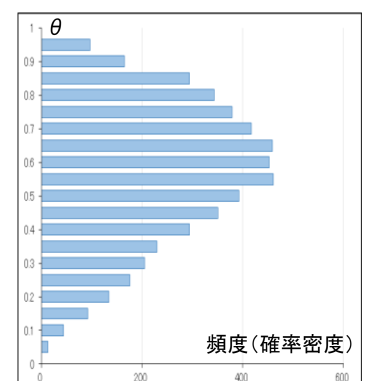
（参考）MCMC の計算過程をビジュアル化したレポートの一例 URL: The Stata Blog » Introduction to Bayesian statistics, part 2: MCMC and the Metropolis–Hastings algorithm



（図13）あぶりだすべき事後分布
（ベータ分布 $f(x) = (1-x)\,x^2$ を設定した）



（図14）マルコフ連鎖の点列と事後分布の値



（図15） （図14）の頻度

12－8. カルマンフィルター（推定）

カルマンフィルターは、例えば車の位置をGPSデータを使用して推定するために使用されている。気象予報のガイダンスなどにも使用されている。動的なシステムの内部状態（例えば、位置、雨量などの変数、状態変数という）を、観測値からノイズを取り除いて（フィルタリングして）推定する方法である。カルマンフィルターの処理手順は、概略以下の通りである。

■処理手順

内部状態は状態変数の動的な時間変化の式（状態方程式）で表し、状態変数を観測する方法は観測方程式で表す。状態変数は確率分布に従うと考え（事前分布）、状態変数を観測データとして入手し、事前分布を更新して事後分布（推定）とする。この事後分布は、状態方程式で示す動的な変化に従って次の時刻の確率分布として扱う。以降、この分布を次の時刻の事前分布とみなして上述の処理を繰り返す。

下記に、正規分布に従う状態変数のカルマンフィルターの式をまとめておいた（導出方法は書籍参照のこと）。

・状態方程式

$x(t) = b\,x(t-1) + m + c\,w$ ・・・（式10）

- $x(t)$： 時刻 t の状態変数、 正規分布で $N(\mu(t),\ \sigma^2(t))$
- $x(t-1)$： 時刻 t－1 の状態変数、正規分布で $N(\mu(t-1),\ \sigma^2(t-1))$
- w： ノイズ、 正規分布で $N(0,\ \rho^2)$
- b, m, c： 定数

・観測方程式

$y(t) = a\,x(t) + v$ ・・・（式11）

- $y(t)$： 時刻 t の観測変数
- v： ノイズ、 正規分布で $N(0,\ \tau^2)$
- a： 定数

■カルマンフィルターの式

・予測の更新ルール（状態方程式）

時刻 t－1 から、時刻 t へ状態遷移

$\mu(t) = E(x(t)) = b\,\mu(t-1) + m$ ・・・（式12）

$\sigma^2(t) = V(x(t)) = b^2\sigma^2(t-1) + c^2\rho^2$ ・・・（式13）

・観測の更新ルール（観測方程式）： 確率分布のパラメータの更新

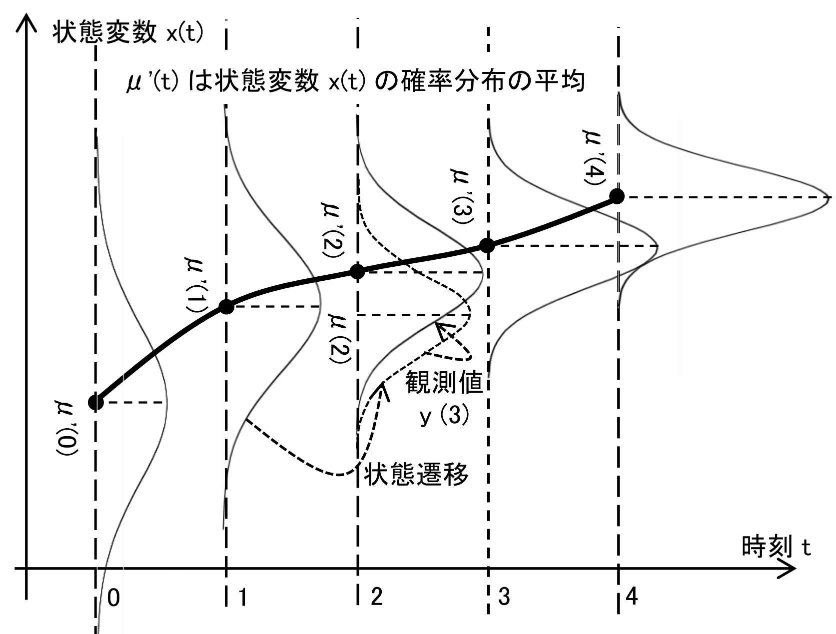
時刻 t で観測データ $y(t)$ を得て、事前分布 $N(\mu(t),\ \sigma 2)$ を事後分布 $N(\mu'(t),\ \sigma'^2(t))$ に更新する

$p(x(t) \mid y(t)) \propto p(y(t) \mid x(t))\,p(x)$ から下式を得る。

$\mu'(t) = (\mu(t)\tau^2 + a\sigma^2(t)\,y(t)) / (\tau^2 + a^2\sigma^2(t))$ ・・・（式14）

$\sigma'^2(t) = \sigma^2(t)\tau^2 / (\tau^2 + a^2\sigma^2(t))$ ・・・（式15）

（図16）は、カルマンフィルターを使った、状態変数の推定の模式図である。時刻とともに確率分布の分散が小さくなっていく様子を表している。
太い実線は推定データの推移を表す。



（図16） カルマンフィルターを使った推定の模式図

12－9. 階層ベイズモデル

データの個性（個体差）を認めるデータ分析を従来の統計学で取り扱うのは難しい。この課題を解決するのが階層ベイズモデルであり、個体差を表すためのパラメータを複数用意することによって、データの持つ個体差にも着目した分析を可能にする。例えば商品広告の効果を評価する場合、全体に共通する効果を調べるだけでなく、地域性とか「個」にも着目した分析を可能にする。

ベイズの基本公式（式8）を拡大して（式14）を用意する。

（式14）の模式図が（図17）であり、母数が階層構造になっているので階層ベイズモデルと呼ばれる。

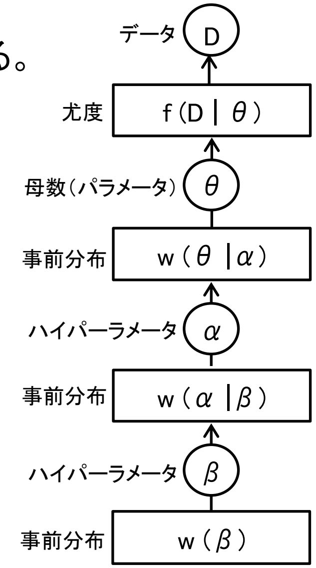
$$\underbrace{w(\theta,\alpha,\beta \mid D)}_{\substack{\text{事後分布}\\\text{（母数の分布）}}} = \underbrace{f(D \mid \theta)}_{\substack{\text{尤度}\\\text{（データの分布）}}}\ \underbrace{w(\theta \mid \alpha)}_{\substack{\text{事前分布}\\\text{（母数}\,\theta\,\text{の分布）}}}\ \underbrace{w(\alpha \mid \beta)}_{\substack{\text{事前分布}\\\text{（母数}\,\alpha\,\text{の分布）}}}\ \underbrace{w(\beta)}_{\substack{\text{事前分布}\\\text{（母数}\,\beta\,\text{の分布）}}} / \underbrace{P(D)}_{\substack{\text{周辺尤度}\\\text{（定数）}}} \quad \cdots \quad \text{（式14）}$$

ここで、D ： データ

θ, α, β ： いずれも、確率分布の母数（平均、分散などのパラメータ）

α は θ の分布を規定する母数（パラメータ）、β は α の分布を規定する母数であり、

α, β は、いずれも ハイパーパラメータと呼ばれる



（図17）階層ベイズモデルの模式図

下記の（式15）は、（式14）をもとに個体差と共通性を加味した階層ベイズモデルである。

$$\overbrace{w(\{a_i\}, b, \alpha, \beta \mid \{D_i\})}^{\substack{\text{事後分布}\\\text{（母数}\,a_i,\,b,\,\alpha,\,\beta\,\text{の分布）}}} \propto \overbrace{\prod_{i=1}^{50} f(D_i \mid a_i, b)}^{\substack{\text{尤度}\\\text{（データの分布）}}}\ \underbrace{\overbrace{w(a_i, \alpha)}^{\substack{\text{事前分布}\\\text{（母数}\,a_i\,\text{の分布）}}}\ \overbrace{w(\alpha)}^{\substack{\text{事前分布}\\\text{（母数}\,\alpha\,\text{の分布）}}}}_{\text{個体差（個性）}}\ \underbrace{\overbrace{w(b)}^{\substack{\text{事前分布}\\\text{（母数 b の分布）}}}}_{\text{共通性（全体）}} \quad \cdots \quad \text{（式15）}$$

ここで、Π は総乗（積）の記号

$\{a_i\} = a_1, a_2, \cdots, a_{50}$, $\{D_i\} = D_1, D_2, \cdots, D_{50}$, $i = 50$ とした

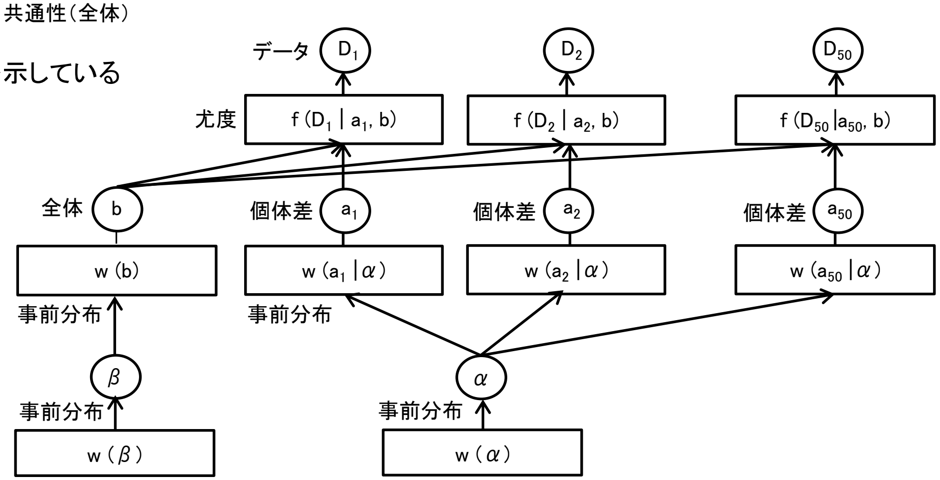
（式7）の定数 P (D) は割愛し、比例関係（$\propto$ ：比例記号）にあることを示している

（図18）は（式15）の模式図である。

ここで、事前分布 $w(a_i \mid \alpha)$ は母数 α の影響が小さければ、同じような分布になり、個体差を示す母数 a_i も同じような値になる。逆に、母数 α の影響が大きければ個体差が大きい統計モデルになる。よって、母数 α は統計モデルの個体差と共通性のバランスをとっているパラメータであると言える。

なお、母数 b は全体に係るパラメータであり、その事前分布を規定する母数 β に制約を特に設ける必要はない。

なお、確率分布 $f(D_i \mid a_i, b)$ の構成に、ロジットモデルという確率分布 $1 / (1 + e^{(-b-a_i)})$ がよく使用されている。



（図18）個体差を活かす階層ベイズモデル

付録1．ベイジアン ネットワークの計算例

自社レストランに関わる4つの変数、ここでは、自社レストランの評判と客数と売上高、そして他社レストランの販促の様子を挙げ、これらの変数の間の因果関係の大きさを調べる。（付表１）が自社レストラン20店舗の４変数のデータ（仮のデータ）である。正規分布に従うデータであるとした。（付表１）を標準化して（付表２）を得る。

（付表１）計算例のためのデータ

変数		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	平均	標準偏差
自社評判	x1	66	60	85	52	59	66	99	99	71	49	69	52	58	45	76	63	103	44	58	76	68	18
他社販促	x2	34	34	26	20	33	30	31	21	25	26	19	34	33	12	7	27	20	20	31	21	25	8
客数	x3	823	826	975	535	744	836	1005	942	955	523	784	646	698	442	912	788	1067	505	655	942	780	183
売上高	x4	1690	1620	1749	1211	1457	1630	1860	1866	1690	1211	1408	1389	1401	1180	1708	1580	1708	1230	1376	1479	1522	218

（付表２）標準化データ

変数		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	平均	標準偏差
自社評判	x1	−0.085	−0.425	0.991	−0.878	−0.481	−0.085	1.784	1.784	0.198	−1.048	0.085	−0.878	−0.538	−1.274	0.481	−0.255	2.010	−1.331	−0.538	0.481	0.0	1.0
他社販促	x2	1.152	1.152	0.105	−0.681	1.021	0.628	0.759	−0.550	−0.026	0.105	−0.811	1.152	1.021	−1.728	−2.382	0.236	−0.681	−0.681	0.759	−0.550	0.0	1.0
客数	x3	0.234	0.251	1.065	−1.340	−0.198	0.305	1.229	0.884	0.955	−1.405	0.021	−0.733	−0.449	−1.848	0.720	0.043	1.567	−1.504	−0.684	0.884	0.0	1.0
売上高	x4	0.770	0.449	1.041	−1.427	−0.299	0.495	1.550	1.577	0.770	−1.427	−0.524	−0.611	−0.556	−1.570	0.853	0.265	0.853	−1.340	−0.670	−0.198	0.0	1.0

（付表２）から分散共分散行列 **S**（付式１）とその逆行列 **P**（付式２）を求める。逆行列はExcel の MINVERSE 関数を使用して求めた。

次いで、分散共分散行列の逆行列 **P** と偏相関行列 **Q** との関係式（付式３）を使って、偏相関行列 **Q**（式４）を求めた。

$$q_{ij} = -p_{ij} / (p_{ii} \cdot p_{jj})^{0.5} \quad \cdots \text{（付式3）}$$

この偏相関係数が、ベイジアンネットワーク（図８）の因果関係の強さを表す。

$$\mathbf{S} = \begin{bmatrix} S_{x1x1} & S_{x1x2} & S_{x1x3} & S_{x1x4} \\ S_{x2x1} & S_{x2x2} & S_{x2x3} & S_{x2x4} \\ S_{x3x1} & S_{x3x2} & S_{x3x3} & S_{x3x4} \\ S_{x4x1} & S_{x4x2} & S_{x4x3} & S_{x4x4} \end{bmatrix} = \begin{bmatrix} 1 & -0.0971 & 0.8564 & 0.8127 \\ -0.0971 & 1 & 0.0482 & 0.1229 \\ 0.8564 & 00482 & 1 & 0.8706 \\ 0.8127 & 0.1229 & 0.8706 & 1 \end{bmatrix} \quad \cdots \text{（付式1）}$$

$$\mathbf{P} = \begin{bmatrix} P_{x1x1} & P_{x1x2} & P_{x1x3} & P_{x1x4} \\ P_{x2x1} & P_{x2x2} & P_{x2x3} & P_{x2x4} \\ P_{x3x1} & P_{x3x2} & P_{x3x3} & P_{x3x4} \\ P_{x4x1} & P_{x4x2} & P_{x4x3} & P_{x4x4} \end{bmatrix} = \begin{bmatrix} 4.517 & 0.748 & -2.596 & -1.503 \\ 0.748 & 1.154 & -0.180 & -0.594 \\ -2.596 & -0.180 & 5.684 & -2.817 \\ -1.503 & -0.594 & -2.827 & 4.746 \end{bmatrix} \quad \cdots \text{（付式2）}$$

$$\mathbf{Q} = \begin{bmatrix} Q_{x1x1} & Q_{x1x2} & Q_{x1x3} & Q_{x1x4} \\ Q_{x2x1} & Q_{x2x2} & Q_{x2x3} & Q_{x2x4} \\ Q_{x3x1} & Q_{x3x2} & Q_{x3x3} & Q_{x3x4} \\ Q_{x4x1} & Q_{x4x2} & Q_{x4x3} & Q_{x4x4} \end{bmatrix} = \begin{bmatrix} -1 & -0.328 & 0.512 & 0.325 \\ -0.328 & -1 & -0.070 & 0.254 \\ 0.512 & -0.070 & -1 & 0.542 \\ 0.325 & 0.254 & 0.542 & -1 \end{bmatrix} \quad \cdots \text{（付式4）}$$

付録2. カルマンフィルターの計算例

カルマンフィルターの（式12）と（式14）、および（式13）と（式15）をそれぞれ一つの式にまとめる。

・平均の更新の式　（式12）と（式14）から

$\mu'(t) = b\mu(t-1) + m + G(y - a(b\mu(t-1) + m)) = ① + ③ \times ②$ ・・・（付式5）

$\mu'(t)$：　時刻 t の確率分布 x(t) の $\mu(t)$ の推定値

① $b\mu(t-1) + m$ ：時刻 t − 1 の確率分布 x (t − 1) の $\mu(t-1)$ から求めた時刻 t の $\mu(t)$（理論推定値であり、ノイズの考慮はなし）

② $y - a(b\mu(t-1) + m)$：　時刻 t − 1 の観測データ y (t) と理論推定値に基づく値との差分

③ $G = a(b^2\sigma^2(t-1) + c^2\rho^2) / (\tau^2 + a^2(b^2\sigma^2(t-1) + c^2\rho^2))$ ・・・（付式6）：　差分を理論推定値に反映する度合いの係数（カルマンゲイン）

・分散の更新の式　（式13）と（式15）から

$\sigma'(t) = (b^2\sigma^2(t-1) + c^2\rho^2)\tau^2 / (\tau^2 + a^2(b^2\sigma^2(t-1) + c^2\rho^2))$ ・・・（付式7）

$\sigma'(t)$：　時刻 t の確率分布 x (t) の $\sigma(t)$の推定値

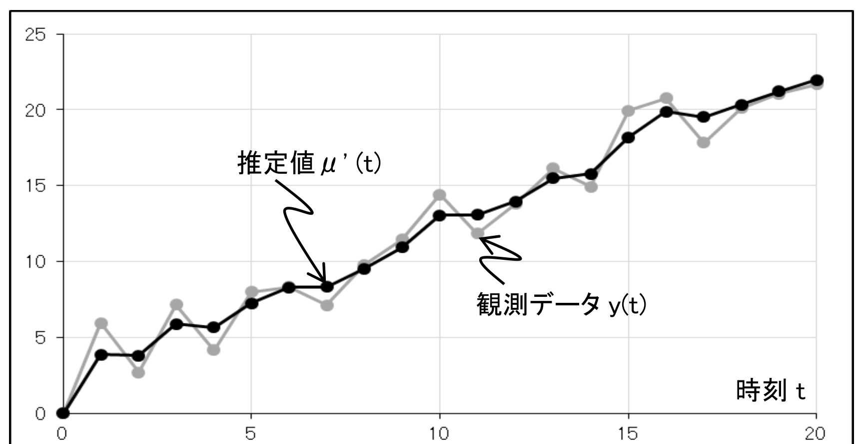
（付式5）は、時刻 t の確率分布 x (t) の $\mu'(t)$ は、①理論推定値に、②観測データとの差分を、③カルマンゲインのウェイトで加味して計算されることを示している。カルマンゲインを 0 にした場合は、観測データを無視して理論推定値を採用することになる。一方、カルマンゲインを 1 にした場合は、観測値 y (t) そのものが推定値になる（推定しないことになる）。カルマンゲインの値によって観測値を加味する度合いを調整できる。なお（付式7）より、$\sigma'(t) \leqq \sigma(t-1)$ であるので時刻の経過とともに確率分布の分散は減少していく。

（付表3）計算例のデータ

t	y(t)	μ'(t)	$\sigma'^2(t)$	G
0	0.000	0.000	1.000	
1	1.681	1.392	0.576	0.576
2	5.754	4.018	0.484	0.484
3	6.673	5.775	0.458	0.458
4	8.721	7.651	0.450	0.450
5	8.420	8.547	0.447	0.447
6	8.948	9.280	0.447	0.447
7	11.585	10.862	0.447	0.447
8	9.132	10.643	0.446	0.446
9	10.908	11.315	0.446	0.446
10	14.338	13.218	0.446	0.446
11	13.806	14.034	0.446	0.446
12	16.447	15.665	0.446	0.446
13	13.670	15.328	0.446	0.446
14	18.737	17.404	0.446	0.446
15	17.186	17.860	0.446	0.446
16	16.509	17.811	0.446	0.446
17	20.744	19.674	0.446	0.446
18	22.228	21.368	0.446	0.446
19	23.240	22.757	0.446	0.446
20	22.657	23.266	0.446	0.446

■計算例

（付式5）と（付式7）で、a = 1, b = 1, c = 0.6, m = 1, $\rho^2 = 1$, $\tau^2 = 1$, $\mu(0) = 0$, $\sigma^2(0) = 1$ と置いて、時刻 t = 0 〜 20 の観測データ y(t) が得られたときの推定値 $\mu'(t)$ を求めた（付図1）。
ここでは、a = 1 としているので、$\mu'(t)$ は y (t) の推定値になる。観測データのノイズを除去して推定している様子、データを平滑化している様子が分かる。
なお、カルマンゲインは 0.446 で、約 45% のウェイトで観測データの情報を活用している。
また、ここでの設定では $\sigma'^2(t)$ とカルマンゲインの値は同じになる。



（付図1）カルマンフィルターを使った推定

参考にした書籍

河本薫：「会社を変える分析の力」、 講談社現代新書
小山昇：「数字は人格」、 ダイヤモンド社
永野裕之：「ビジネス×数学＝最強」、 すばる舎
竹内薫：「数学×思考＝ざっくりといかにして問題をとくか」、 丸善出版
中西達夫：「統計データをすぐに分析できる本」、 アニモ出版
中西達夫：「すぐれた判断が統計データ分析から生まれる」、 実務教育出版
豊田裕貴：「マンガでわかる　ビジネスを成功に導くデータ分析」、 ナツメ社
向後千春、冨永敦子：「統計学がわかる」、 技術評論社
石井俊全：「意味がわかる統計学」、 ベレ出版
涌井良幸、涌井貞美：「中学数学でわかる統計の授業」、 日本実業出版
涌井良幸、涌井貞美：「統計学の図鑑」、 技術評論社
西内啓：「統計学が最強の学問である」、 ダイヤモンド社
西内啓：「統計学が最強の学問である（実践編）」、 ダイヤモンド社
西内啓：「統計学が最強の学問である（ビジネス編）」、 ダイヤモンド社
森岡毅、今西聖貴：「確率思考の戦略論」、 角川書店
デビッド・マクアダムス：「世界一流企業はゲーム理論で決めている」、 ダイヤモンド社
河村真一ほか：「本物のデータ分析力が身に付く本」、 日経BPムック
末吉正成、末吉美貴：「Excel　ビジネス統計分析　この分析できますか？」、 翔泳社
谷岡一郎：「社会調査のウソ　リサーチ・リテラシーのすすめ」、 文藝春秋
林知己夫：「調査の科学」、 ちくま学芸文庫
八谷大岳：「データ解析」シリーズ　全１５回、 （例）データ解析 第1回 ベクトルの復習 – YouTube
涌井良幸、涌井貞美：「身につくベイズ統計学」、 技術評論社
松原望：「よくわかる最新ベイズ統計の基本と仕組み」、 秀和システム
藤田一弥：「見えないものをさぐる　それがベイズ　～ツールによる実践ベイズ統計～」、 オーム社
久保田拓弥：「データ解析のための統計モデリング入門 – 一般線形モデル・階層ベイズモデル・MCMC」、 岩波書店
久保田拓弥：「階層ベイズ＆MCMC講義」、 階層ベイズ＆MCMC講義（久保拓弥） 難易度★★ - YouTube
山本直樹：「ベイズの方法　カルマンフィルター」、慶應大学講義 応用確率論 第十三回　ベイズの方法 カルマンフィルタ1 – YouTube
山本直樹：「ベイズの方法　カルマンフィルター」、慶應大学講義 応用確率論 第十四回　カルマンフィルター2 - YouTube