

法と統計学 第8回 ベイズの定理(4)

森 大輔

熊本大学法学部

2017年

本日の流れ

- 1 ベイズの定理の応用 2
- 2 ベイズの定理に関する有名な問題 2
- 3 ベイズ更新と逐次性

復習：コリンズ事件

- カリフォルニア州である夫人が何者かに突き飛ばされ財布を奪われた
 - 立ち上がると、走り去ってゆく金髪の若い女性の後ろ姿を目撃した
- 目撃者が同時刻に次のことを目撃
 - 叫び声のする辺りから走ってきた女性が、路上に駐車した黄色の車の助手席に飛び乗った
 - 口ひげとあごひげを生やした黒人男性が運転しているのが見えた
 - 助手席の女性は、金髪をポニーテールにした中肉中背の白人女性だった
- 検察が被害者や目撃者の証言に合う人物として、コリンズ夫妻を逮捕
 - 検察は、裁判で立証に困って、大学の数学講師を呼んだ

コリンズ事件

- 自動車を所有している男女のカップルが次に述べる6つの特徴を全て満たす確率
 - 1 車の色は黄色である： $1/10$
 - 2 異人種間カップルである： $1/1000$
 - 3 男性はあごひげを生やした黒人である： $1/10$
 - 4 男性は口ひげを生やしている： $1/4$
 - 5 女性は金髪である： $1/3$
 - 6 女性はポニーテールにしている： $1/10$
- 各特徴が独立だとすれば、確率の積の法則から $1/1200$ 万
 - 各特徴が本当に独立といえるかという問題があった
- さらに、各特徴が本当に独立だとしても、まだ「検察官の誤り」の問題

コリンズ事件

- 「検察官の誤り」
 - 検察は、コリンズ夫妻が犯人でない確率を $1/1200$ 万であるかのように述べた
 - 言い換えれば、検察は、コリンズ夫妻が犯人である確率を $(1 - 1/1200 \text{ 万}) \times 100 = 99.999 \dots \%$ であるかのように述べた
 - 陪審もおそらくそれを信じた
 - コリンズ夫妻は有罪となった
 - しかしこれは誤り
- $1/1200$ 万は、カリフォルニア州のカップルが、犯人であろうがなかろうが、6つの特徴を全て持つ確率

※補足：フェルミ推定

- 数学の手法ではなく、思考法の一つ
- 「米国のシカゴ市には何人のピアノ調律師がいるか？」
- 下のように考えれば大雑把にでも理屈のある値を回答可
 - まず、次のように細かく仮定する
 - シカゴ市の人口は、300万人とする
 - 1世帯あたりの人数が平均3人程度とする
 - 10世帯に1台の割合で、ピアノを保有しているとする
 - ピアノ1台の調律は、平均して1年に1回行うとする
 - 調律師が1日に調律するピアノの台数は、3台とする
 - 週休二日とし、調律師は年間に約250日働くとする
 - 次に、これを使って次のように推論する
 - シカゴ市の世帯数は $300 \text{ 万} / 3 = 100 \text{ 万}$ 世帯程度
 - ピアノの総数は $100 \text{ 万} / 10 = 10 \text{ 万}$ 台程度
 - ピアノの調律は、年間に10万件程度
 - 1人の調律師は1年間に $250 \times 3 = 750$ 台程度を調律
 - よって調律師の人数は $10 \text{ 万} / 750 \approx 130$ 人程度

練習問題 6 : コリンズ事件

カリフォルニア州でひったくり事件が起こったカリフォルニア州にはカップルは 2400 万組おり、この中のどれかが犯人であることがわかっている。

この 2400 万組の中のあるカップルが逮捕された。このカップルは、目撃証言から導き出した、犯人カップルの 6 つの特徴を全て持っている (証言に誤りはない)。

検察によれば、カリフォルニア州のカップル (犯人も含む) が、6 つの特徴を全て持つ確率は $1/1200$ 万である。

- (1) カリフォルニア州には 6 つの特徴を全て持つカップルは何組いるか？
- (2) このカップルが犯人である確率はいくらか？

練習問題 7 : 三囚人問題

3 人の囚人 A、B、C がいる。1 人が恩赦で釈放され、残り 2 人が処刑されることがわかっている。誰が恩赦になるかはまったくわからず、確率は 3 人とも等しいとする。

誰が釈放されるか知っている看守に対し、A が「B と C のうち少なくとも 1 人処刑されるのは確実なのだから、2 人のうち処刑される 1 人の名前を教えてくださいも私についての情報を与えることにはならない。だから処刑される 1 人を教えてくださいくれないか」と頼んだ。看守は A の言い分に納得し「B は処刑される」と答えた。

この答えを聞いた後、A が釈放される確率はいくらになるか。

練習問題 8 : 変形三囚人問題

3 人の囚人 A、B、C がいる。1 人が恩赦になって釈放され、残りの 2 人が処刑されることがわかっている。恩赦になる確率はそれぞれ $1/4$ 、 $1/4$ 、 $1/2$ であるとする。

誰が釈放されるか知っている看守に対し、A が「B と C のうち少なくとも 1 人処刑されるのは確実なのだから、2 人のうち処刑される 1 人の名前を教えてくれても私についての情報を与えることにはならない。だから処刑される 1 人を教えてくれないか」と頼んだ。看守は A の言い分に納得し「B は処刑される」と答えた。

この答えを聞いた後、A が釈放される確率はいくらになるか。

三囚人問題と変形三囚人問題

- 三囚人問題は、数学者で数学の解説の本などを多数書いたマーティン・ガードナーによって紹介
- その後、心理学の分野などで研究
 - 正しい答えを導けないのはなぜか
 - 正しい答えを示されても納得できない人が多いのはなぜか
 - 三囚人問題の正しい答えでいったん納得しても、変形三囚人問題で再び悩み出す人がいるのはなぜか
- モンティ・ホール問題と似ている
- ただ、結論は少し違う
 - 三囚人問題では看守の答えの後、確率は変わらない
 - 変形三囚人問題では看守の答えの後、確率は下がる

練習問題 3: ナイーブベイズフィルタ 2

ある人は、送られてくるメールのうち何%が迷惑メールかわからないので、とりあえず 50%と考えている。

過去のメールから、迷惑メールに「激安」という言葉が含まれる確率は 60%である。また、迷惑メールでないものに「激安」という言葉が含まれる確率は 20%である。

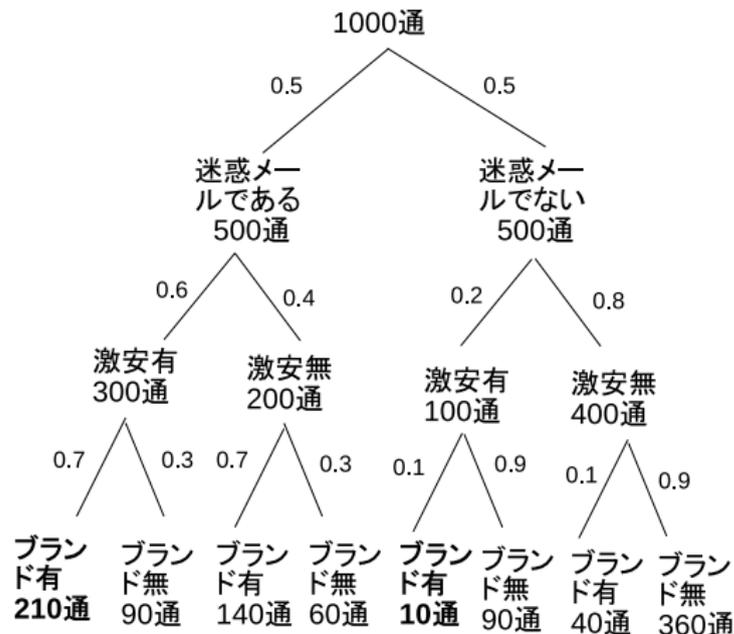
さらに、過去のメールから、迷惑メールに「ブランド」という言葉が含まれる確率は 70%である。また、迷惑メールでないものに「ブランド」という言葉が含まれる確率は 10%である。いま、この人に届いた 1 通のメールに、「激安」と「ブランド」という 2 つの言葉が両方含まれていた。

この場合、このメールが迷惑メールである確率はいくらと考えられるか。ただし、「激安」が含まれることと「ブランド」が含まれることは、独立だと考えること。

ナイーブベイズフィルタ 2

- 今回の問題は前回のものより現実のフィルタに近い
 - 例えば、迷惑メールである確率が95%より大きい場合に、迷惑メールフォルダに振り分ける設定とする
 - 確率は計算すると95.5%程度となるため、今回のメールは迷惑メールフォルダに
- 「ナイーブ」とは単純という意味
 - ある語が含まれることと別の語が含まれることを、独立だと考えている点が単純
 - 実際は独立ではないことが多い
 - 「激安」と「ブランド」は同じ文脈で出てきそう
 - しかし、独立と考えることで情報が少なくてすむ
 - 独立でないと、「激安」があるときに「ブランド」がある確率と、「激安」がないときに「ブランド」がある確率を別々に知る必要
 - しかも実際に意外とうまく迷惑メールを振り分けられる

独立性の樹形図上での表現



- 「迷惑メールか否か」の枝の下に「激安有か無か」の枝、さらにその下に「ブランド有か無か」の枝を書く
- 左の図から、求める確率は

$$210 / (210 + 10) \doteq 0.955$$

独立性の樹形図上での表現

■ 理由不十分の原理

- 「ある人は、過去1年間に送られてきたメールのうち迷惑メールが何%か分からないので、とりあえず50%が迷惑メールと考えている。」
- このように、樹形図の一番上の枝の確率が分からないときは、とりあえず等確率に割り振っておくことがよく行われる
 - これを理由不十分の原理と呼ぶ

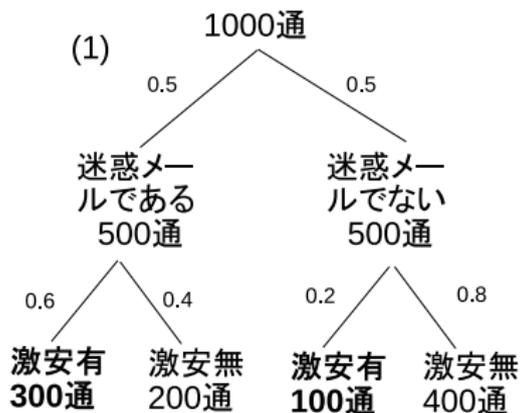
■ 独立

- 「ブランド」が含まれるか否かは、「激安」が含まれるか否かに影響されない、ということ
 - 「激安」が含まれていようがまいが、迷惑メールに「ブランド」が含まれる確率は70%
 - 「激安」が含まれていようがまいが、迷惑メールでないものに「ブランド」が含まれる確率は10%

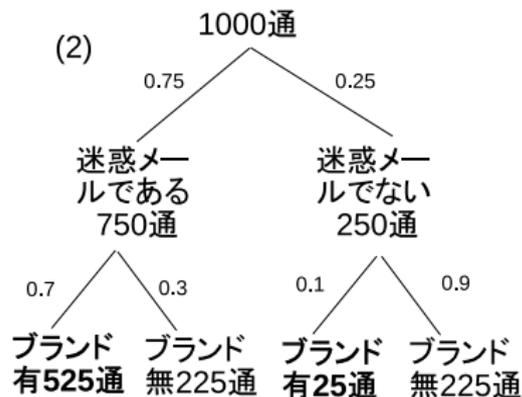
逐次性

- この問題は、次のように2つに分けて考えても同じ答えが出る（逐次性）
 - (1) いま、届いた1通のメールに、「激安」という言葉が含まれていた。このメールが迷惑メールである確率は？
 - メールのうち50%が迷惑メール
 - 過去のメールから、迷惑メールに「激安」という言葉が含まれる確率は60%で、迷惑メールでないものに「激安」という言葉が含まれる確率は20%
 - (2) さらに、(1)のメールには、「ブランド」という言葉も含まれていた。このメールが迷惑メールである確率は？
 - メールのうち(1)で計算した確率のものが迷惑メール
 - 過去のメールから、迷惑メールに「ブランド」という言葉が含まれる確率は70%で、迷惑メールでないものに「ブランド」という言葉が含まれる確率は10%

逐次性



- $300 / (300 + 100) \doteq 0.75$
 - これを今度は右の樹形図の一番上の枝で使う

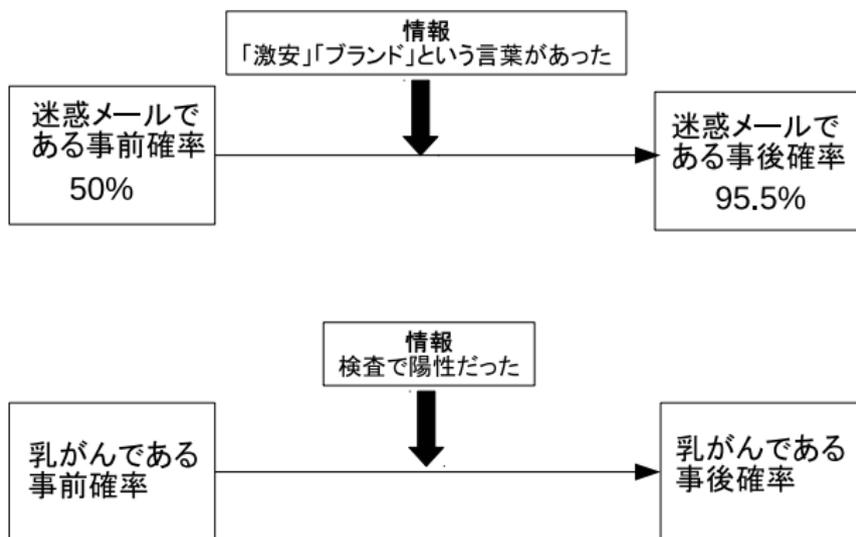


- $525 / (525 + 25) \doteq 0.955$
 - これは一気に2つ考えた時の数字と同じ

ベイズ更新

- 樹形図の一番上の枝分かれのところの確率を事前確率と呼ぶ
 - 迷惑メールの問題では、送られてきたメールのうち迷惑メールが何%かということ
 - わからない場合は理由不十分の原理で50%としていた
- 今までの問題で求めていた確率を事後確率と呼ぶ
 - 迷惑メールの問題では、送られてきたメールに「激安」「ブランド」という言葉があるとき、それが迷惑メールである確率
- 今までの問題は、事前確率が最初にあって、そこに情報が加わったとき、事後確率がどうなるか、と見ることを考えてきたといえる
 - 事前確率が、得た情報によって事後確率に更新される
 - これをベイズ更新という

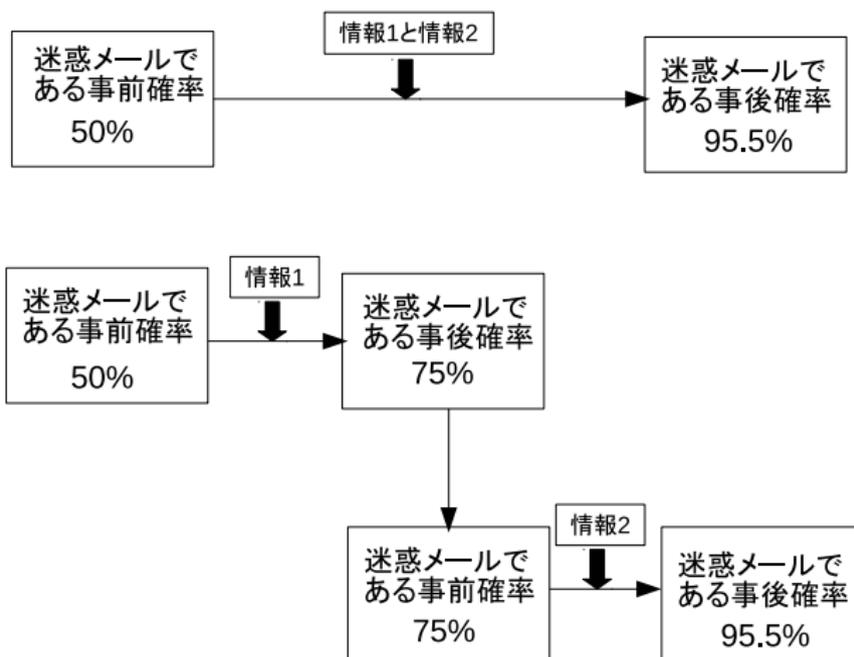
ベイズ更新



ベイズ更新と逐次性

- ベイズ更新は、まとめて行っても逐次的に行っても、同じ結果となる
 - これを逐次性と呼ぶ
 - つまり、情報が入ってくるたびに、確率が更新されているプロセスだと見ることができる
 - これは非常に重要な見方
- ある種の学習機能と見ることもできる
 - 情報を取り入れて、賢くなる
- 逐次性は、確率が独立でもそうでなくても、どちらでも成り立つ
 - 今のナイーブベイズフィルタの問題では、確率は独立だとしていた

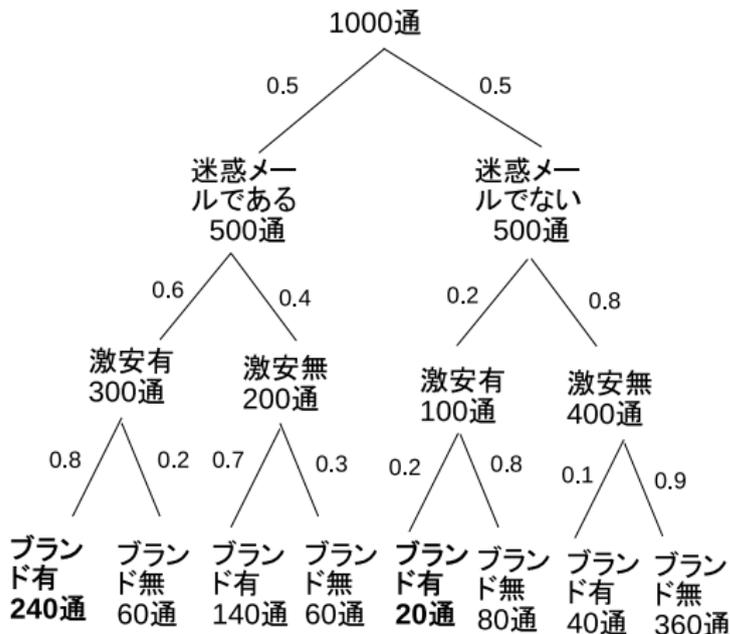
ベイズ更新と逐次性



独立性と逐次性

- 先ほどの迷惑メールの問題
 - 以下はそのまま
 - 迷惑メールである事前確率は 50%
 - 迷惑メールに「激安」がある確率は 60%で、迷惑メールでないものに「激安」がある確率は 20%である。
 - 以下を変更
 - 独立性を仮定しない
 - 迷惑メールに「激安」があるときに、「ブランド」がある確率は 80%
 - 迷惑メールに「激安」がないときに、「ブランド」がある確率は 70%
 - 迷惑メールでないものに「激安」があるときに、「ブランド」がある確率は 20%
 - 迷惑メールでないものに「激安」がないときに、「ブランド」がある確率は 10%
 - 一気に解く場合と逐次的に解く場合を比較

一気に解く場合

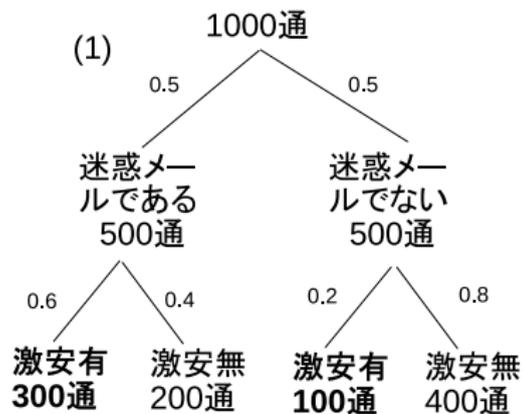


- 独立でない場合、一番下の枝分かれ4つの確率をそれぞれ考える必要がある

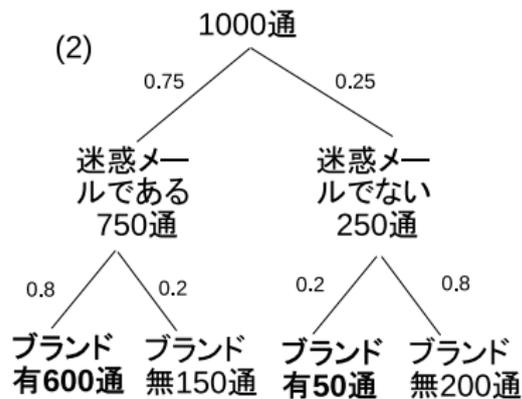
- 独立の場合は、枝分かれ2つのみわかればよい

- 左の図から、求める確率は $240 / (240 + 20) \doteq 0.923$

逐次的に解く場合



■ $300 / (300 + 100) \doteq 0.75$



■ $600 / (600 + 50) \doteq 0.923$

独立性と逐次性

- 逐次的に解く場合
 - 前のスライドの (1) の部分は、前の問題とまったく同じ
 - (2) の部分の下の枝分かれの確率は、「激安」がある場合のものを使う
 - 左の枝分かれでは、「迷惑メールに「激安」があるときに、「ブランド」がある確率は 80%」を使う
 - 右の枝分かれでは、「迷惑メールでないものに「激安」があるときに、「ブランド」がある確率は 20%」を使う
 - メールに「激安」があることは既にわかっているので、これらを使う
- このように、確率が独立であろうとなかろうと、ベイズ更新の逐次性という性質は成り立つ
- これは非常に便利な性質
 - 情報が最初から全部そろっている必要はなく、情報が入るたびに、確率を更新していけばよい

参考文献

- マイクル O. フィンケルスタイン (2014) 『法統計学入門：法律家のための確率統計の初歩』 木鐸社の第1章.
- ゲルト・ギーゲレンツァー (2010) 『リスク・リテラシーが身につく統計的思考法：初歩からベイズ推定まで』 早川書房.
- 草野耕一 (2014) 「『検察官の誤謬』と『弁護人の誤謬』—ベイズの定理の実践的活用法」 論究ジュリスト 10、20–29.
- 市川伸一 『考えることの科学』 中公新書.
- 田中規久雄 (2013) 「米国における統計的証拠の意義と技法—もう一つの法情報学—」 名古屋大学法政論集 250、347–371.
- 細谷功 (2007) 『地頭力を鍛える：問題解決に活かす「フェルミ推定」』 東洋経済新報社.
- 小島寛之 (2015) 『完全独習 ベイズ統計学入門』 ダイヤモンド社.