Keio University SFC 2004

『探索的モデリング』

第3回 ベイズ確率 reloaded

いば　たかし
井庭 崇
慶應義塾大学総合政策学部 専任講師
iba@sfc.keio.ac.jp
http://www.sfc.keio.ac.jp/~iba/lecture/

天気予報の「降水確率」

- 「天気予報は、よくはずれる」という日常感覚
- 「明日の降水確率は40%」の意味は？

不確実性のモデル

- 不確実性を考える上で重要な「情報」
- 不確実性の源泉
  - 「知識の不足」
  - 「未来の時間」
- 「今日の夕飯はカレーかハンバーグか？」
- 「明日の天気は晴れか雨か？」

血液型の例：不確実性と情報

- 今、あなたが話している相手の血液型を知りたいと思っています。
- 結果を聞く前は、相手の血液型はあなたにとって不確実現象ですが、これは「知識の不足」による不確実性です。
- このとき、日本人全体の血液型の分布を割り当てるのが自然。
- {A,O,B,AB}={4:3:2:1}
- 相手がA型である確率P(A)=0.4

血液型の例：不確実性と情報

- 相手が「自分はB型ではない」と教えてくれたとすると・・・
- 不確実性モデルは変化する。
- B型ではない場合
- {A,O,AB}={4:3:1}
- そうすると、確率も変わる
- P(A | B型ではない) = 4 / (4+3+1) =0.5
- 情報が得られたために、A型である確率は、0.4から0.5に変化している（「足し合わせると1」という性質を保持したため）。
- 情報によって、確率の見積もりは変化する！

コインの判定

- コイン投げをして、本物のコインか、偽者のコインかを判定
- 本物のコインは、投げると表が5割、裏が5割程度出る。
- 偽物のコインは、重さにわずかな偏りがあり、投げると表が6割、裏が4割程度出る。
- とりあえず100回投げてみると、表は49回出た。
- このコインは、本物か偽物か？

コインの判定：フィッシャー流アプローチ

- 偽物のコインを100回投げて49回表が出る確率を計算すると
  - P(表)×P(裏)×P(表)×P(裏)×・・・
  - 0.6×0.4×0.6×0.4×・・・=0.6^49×0.4^51

∧

- 本物のコインを100回投げて49回表が出る確率を計算すると
  - P(表)×P(裏)×P(表)×P(裏)×・・・
  - 0.5×0.5×0.5×0.5×・・・=0.5^49×0.5^51

さいゆう

最尤思想

- 現実におきた現象を説明するモデルが複数あるとき、各モデルでその現象が起きる可能性を計算し、それがもっとも大きくなるモデルを採用する、あるいは確率が著しく小さなモデルは捨てる、そういう思想。
- 「現実に起きているできごとは、もっとも起きやすいできごとである」

コインの判定：ベイズ流アプローチ

- フィッシャー流では、偽物である場合と、本物である場合を場合分けして考えるが、ベイズ流では、一緒に考える。
- 4つの可能性
  - 偽物で表がでる
  - 偽物で裏がでる
  - 本物で表がでる
  - 本物で裏がでる
- 偽物である確率と、本物である確率はわからないので、とりあえず0.5:0.5と割り振っておく。

コインの判定：ベイズ流アプローチ

- 偽物は、投げると表が6割、裏が4割
- 本物は、投げると表が5割、裏が5割

	偽物	本物
表がでる	0.5×0.6	0.5×0.5
裏がでる	0.5×0.4	0.5×0.5

⇒

偽物	本物
6/11	5/11

(0.5×0.6 : 0.5×0.5) = 6:5

「1回投げたら表だった」という情報が得られたので、黄色部分に制限されている。

コインの判定：ベイズ流アプローチ

- 「おもてが出た」という情報を得たあとでのF,Tの確率だと見なせる。
  - P(偽物｜1回目が表)＝ 6/11
  - P(本物｜1回目が表)＝ 5/11
- 偽物である可能性が、情報（データ）を得たことで、5分5分から、6:5へと、やや上昇している。
- 情報（データ）の入手によって、偽物か本物かに関する推測が改定（アップデート）された」ことを表している。

オオカミ少年モデル：ベイズ推定

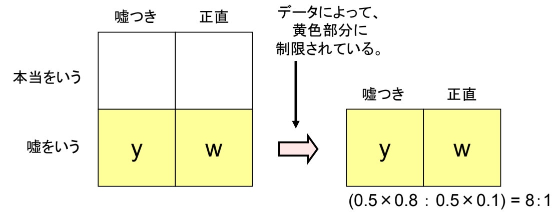
- ある人が信頼できるかどうかの判定
  - 日常的には、まず適当な先入観をもち、その人の行動を観察して、その先入観を修正していく。
- 確率で考えて見ると
  - その人は「嘘つき」か「正直」のどちらか
  - それぞれのケースについて、「嘘をつく確率」を設定する。
    - 「嘘つき」＝「本当のことを0.2の確率でいい、0.8の確率で嘘をつく」
    - 「正直」＝「本当のことを0.9の確率でいい、0.1の確率で嘘をつく」
- その人が1回嘘をついたとすると、その人はどの程度「嘘つき」であるか？

オオカミ少年モデル：ベイズ推定

- まず、その人が「嘘つき」か「正直」かについて、適当に先入観をつくります。
  - 嘘つき：正直＝1：1
  - P(嘘つき)=0.5, P(正直)=0.5
- 4つの可能性
  - 「嘘つき」かつ「本当のことを言う」
  - 「嘘つき」かつ「嘘を言う」
  - 「正直」かつ「本当のことを言う」
  - 「正直」かつ「嘘を言う」

オオカミ少年モデル：ベイズ推定

- 「その人が1回嘘をついた」という情報を入手したので、条件付確率によって確率を変化させることができる。



条件付確率
P(y | 嘘をついた) = 8/9　P(w | 嘘をついた) = 1/9

オオカミ少年モデル：ベイズ推定

- 図をみるとわかるように、嘘をついていることがわかっているy、wへの制限は、「嘘つき」「正直」の区別を表している。
- つまり、先の結果は、次のことを表している。
  - P(嘘つき | 嘘をついた) ＝8/9
  - P(正直 | 嘘をついた) =1/9
- 以上が、1回嘘をついたという情報をもとに、その人が嘘つきかどうかの可能性を見積もる方法
- このプロセスをベイズ推定という。

オオカミ少年モデル：ベイズ推定

- オオカミ少年モデルで得られた確率 8/9 と1/9 を、「ベイズ逆確率」という。
- 「逆」の意味
  - 普通は、「原因から結果へ」という形式
    - 「嘘つき→嘘を言う」という確率P(嘘をつく | 嘘つき)
  - ここで求めたものは、「結果から原因へ」形式
    - 「嘘をついた→嘘つきだ」という確率P(嘘つき | 嘘をつく)

練習問題：ガン検診

- あるガン検診の方法は、ガンの人は0.95の確率でガン陽性と診断され、健康な人がガン陽性と誤診される確率は0.05だとする。（つまり、どちらにしても的中率95%）
- ガンにかかる率を0.005とする。
- この検査の結果、「ガン陽性である」という診断がでたら、自分のガンの可能性をどの程度と疑うべきだろうか？
- 4つの可能性
  - ガンで陽性
  - ガンで陰性
  - 健康で陽性
  - 健康で陰性

ベイズ推定の利点

- 多数回の試行のあとで行ったベイズ推定は正確に真実を見つけ出す。
  - 多数決するなら、フィッシャー流と変わりない結論を下せる。
- 繰り返してベイズ推定を行う場合、いちいち今までの情報をすべて洗いなおして計算するのではなく、最新の情報だけでアップデートすれば結果的に同じ数値をもとづけて導けるという利便性
  - 「逐次合理性」

以上の2つの利点が、理論家だけでなく、エンジニアにも受け入れられた理由

コインの判定：ベイズ流アプローチ

- コインを2回続けて投げて、二回とも表であったら、推定結果がどうなるか？
- 「偽物のコインから表、表、と出る」確率
  - P(F＆表＆表)＝0.5×0.6×0.6=0.18
- 「本物のコインから表、表と出る」確率
  - P(T＆表＆表)＝0.5×0.5×0.5=0.125
- この比をとると
- 0.18 : 0.125 = 36 : 25

コインの判定：ベイズ流アプローチ

- 二回連続で表が出たとき、コインが偽物である、という推定値は、
  - P(F | 表＆表)= 36 / 61
- コインを投げる前の推定値は0.5
- 一回表が出たときの推定値は6/11=約0.55
- 二回表が出たときの推定値は 36 / 61=約0.59
- だんだん偽物のコインだと疑う方へ

コインの判定：ベイズ流アプローチ

- 1回目に表が出たことで、偽物コイン、本物のコインであるという先入観を6：5に変更し、そして一回おもてが出たというデータを忘れてしまったとしましょう。
- 二回目におもてが出たのを観測したあなたは、「6：5という先入観から、表を観測した」として、先入観の改定をすることになります。

ベイズ推定の利点：逐次合理性

- 最初の先入観からデータ二つによってアップデートされた推定値と、データ一つからまず推定値をアップデートし、その推定値を次のデータからアップデートしても、結局は同じ結果になる。
- この性質のおかげでベイズ推定を使って何かを制約するときに、データが加わるたびに全データを洗い直す必要はなく、現在の推定値を新しいデータだけから改訂すればいいので、臨機応変に対応できる。

参考文献

- 『確率的発想法：数学を日常に活かす』（小島 寛之, NHK出版, 2004）