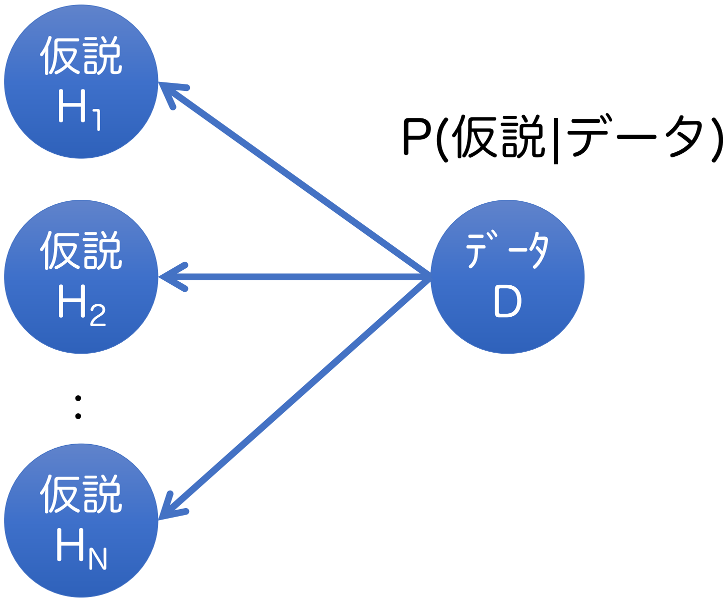
ベイズ統計

古谷知之

講義概要

- ベイズの定理
- ベイズ的意思決定
- ベイズ更新
- 理由不十分の原則
- 事前情報の有無による違いとは

データから仮説（モデル）を推定する



データから仮説（モデル）を推定する

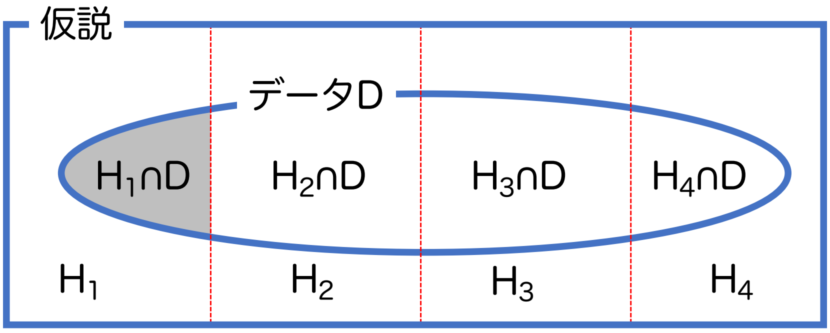
- データDから仮説H_1が成立する確率は以下のようになる

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{\sum_{i=1}^{n} P(D \cap H_i)}$$

$$= \frac{P(D|H_1)P(H_1)}{\sum_{i=1}^{n} P(D|H_i)P(H_i)}$$

$$= \frac{P(D|H_1)P(H_1)}{P(D)}$$

データから仮説（モデル）を推定する



ベイズの定理とベイズ統計

- ベイズ統計でのモデル推定は、データが与えられた条件の下で仮定されたモデルの成立する確率**P(仮説|データ)**を求めていることに他ならない
- 確率的には与えられたデータのもとで様々なモデル（仮説）が成立する可能性がある

尤度（ゆうど）

- 原因から結果が生じる確率P(結果|原因)を尤度という
- 原因と結果を、データと仮説（モデル）と読み替えると、

ある仮説（モデル）を所与としてデータが得られる確率
P(データ|仮説)といえる

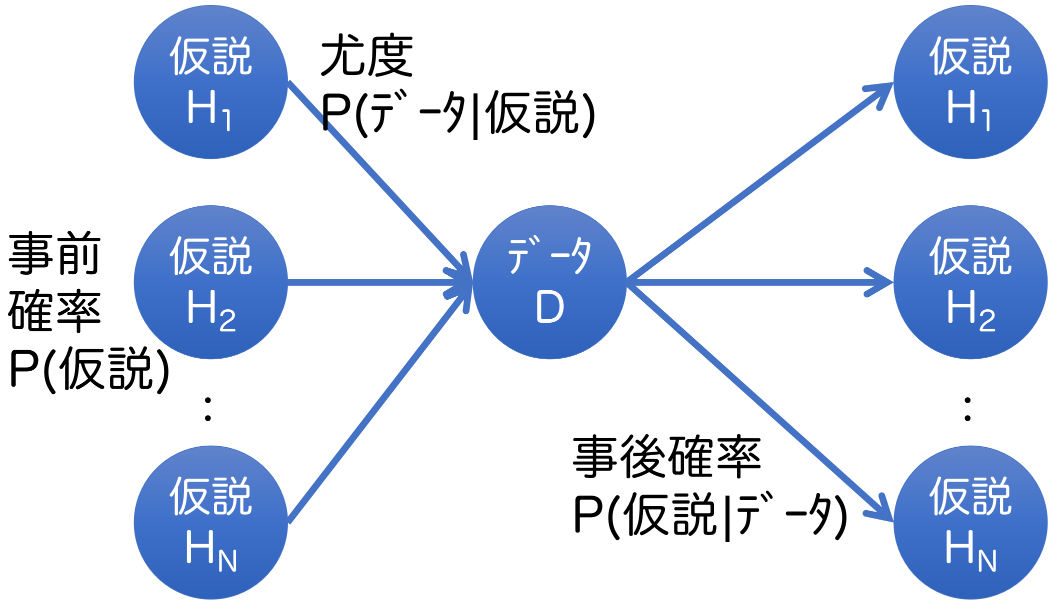
- 仮説がデータに当てはまる当てはまりやすさ（尤もらしさ）を尤度という

事前確率・事後確率・尤度

- ベイズの定理にもとづいて、 $P(H)$を事前確率、 $P(D|H)$を尤度、 $P(H|D)$を事後確率という

$$\overset{\text{事後確率}}{P(H|D)} = \frac{\overset{\text{尤度}}{P(D|H)}\,\overset{\text{事前確率}}{P(H)}}{P(D)}$$

事前確率・尤度・事後確率



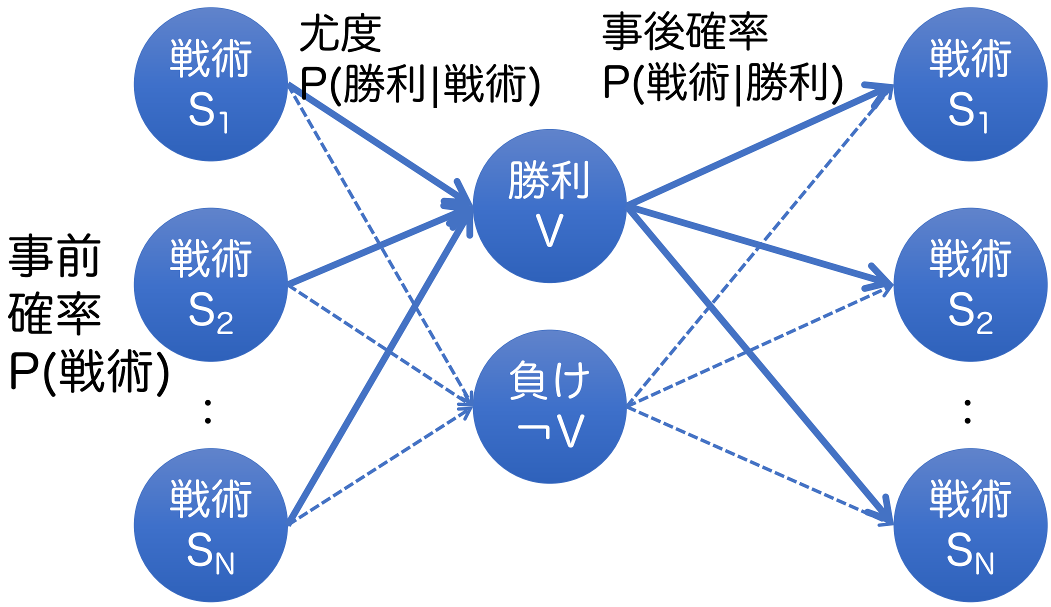
スポーツとベイズ的意思決定の例

- 野球
  - 投手（捕手）の打者に対する配球
  - 打者の投手に対する配球読み
- サッカー・ラグビー・アメフト・バレー
  - 試合の流れの中での戦術の選択
  - パスコース（パスを出す相手）の選択
- ゴルフ
  - クラブの番手選択
  - スイング・パットの強さ・方向選択

『敵を知り己を知れば百戦危うからず』（孫子）

- 今あなたは、あるスポーツチームのアナリスト兼作戦立案者である。次に試合をする相手チームとは初めての対戦であるが、そこでの試合の戦い方を検討中である。
- まずあなたは、相手チームが選ぶであろう戦術を考える＝「敵を知る」
- 次に、相手チームの戦術に応じて自チームの戦術を選ぶ＝「己を知る」

相手チームの戦術の検討



相手チームの選ぶ戦術

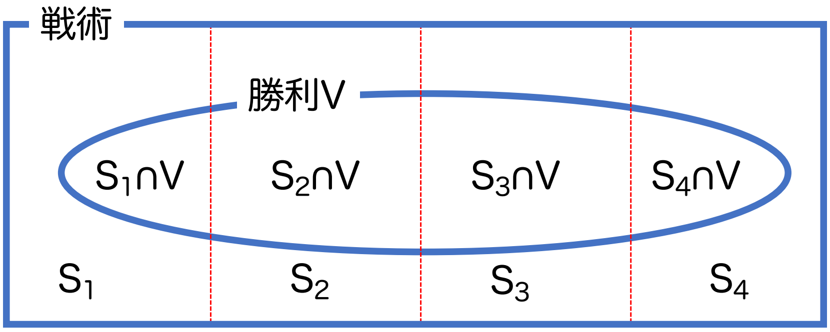
- 過去の対戦結果（勝利した結果V）に基づき、相手チームが勝つという条件付きで戦術S_1を選ぶ事後確率は？

$$P(S_1|V) = \frac{P(V|S_1)P(S_1)}{P(V)}$$

- 相手の戦術選択確率（事前確率）、各戦術の下での勝率（尤度）、相手の勝率がわかれば事後確率がわかる

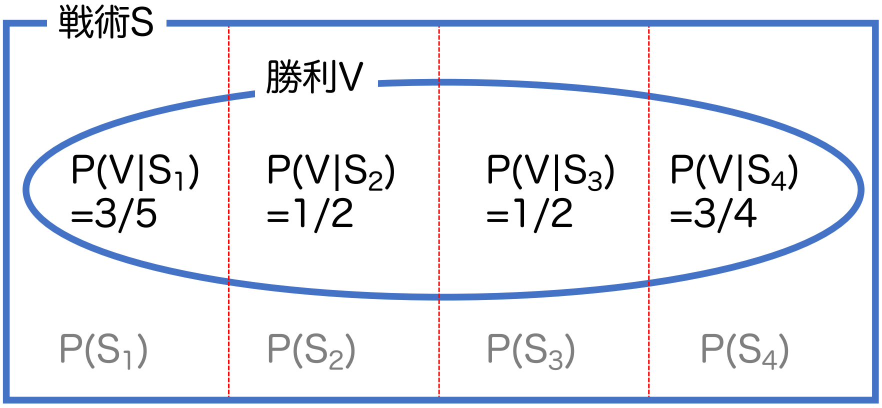
戦術の検討

- この競技では主に４つの戦術（戦術S_1、戦術S_2、戦術S_3、戦術S_4）が試合に使われている。



一般的な戦術選択確率

- 過去の試合などで各チームが選んだ戦術と勝敗との関係を知っているとき



相手チームの戦術の検討

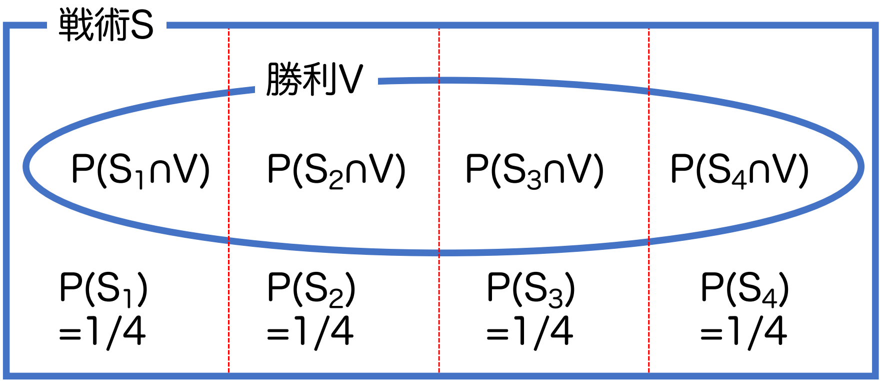
以下のような状況下で、どのように相手チームの戦術を検討すればよいだろうか？

1) 相手チームの戦術に関する情報が全く無いが、相手チームの勝率や一般的に各戦術が選ばれる確率は把握しているとき

2) 過去の対戦経験から相手チームの戦術と勝率との関係を把握しているとき

相手チームの情報がないとき

- 相手の戦術選択確率を等しいと考える
  ＝「理由不十分の原則」という



相手チームの勝率

- 各戦術を選択した条件下での勝率（尤度）と戦術選択率の合計＝勝率

$$P(V)=\sum_{i=1}^{N}P(V|S_i)P(S_i)=\frac{3}{5}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{1}{4}+\frac{3}{4}\cdot\frac{1}{4}$$
$$=0.5875$$

相手チームの情報がないとき

- 「理由不十分の原則」に基づく事後確率
  ⇒$P(S_1) = P(S_2) = P(S_3) = P(S_4) = 1/4$

$$P(S_1|V) = \frac{P(V|S_1)P(S_1)}{P(V)} = \frac{(3/5)\cdot(1/4)}{0.5875} \approx 0.255$$

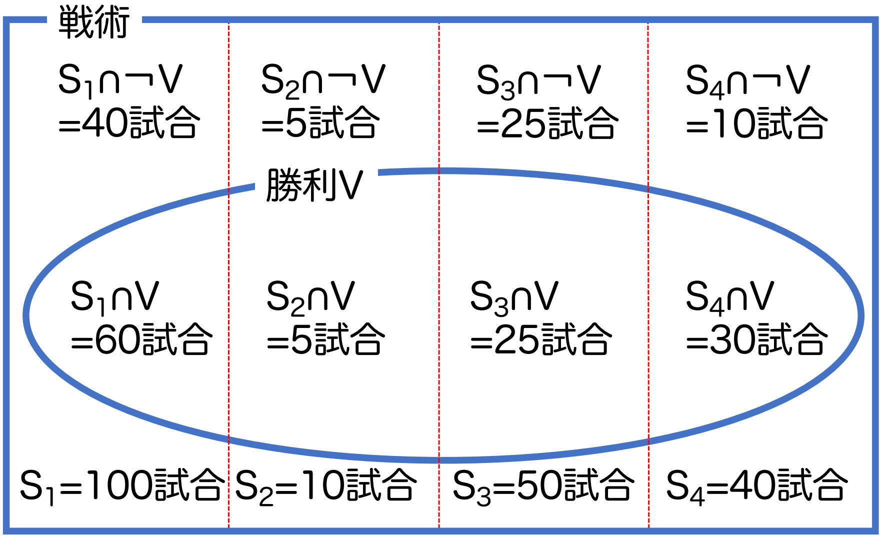
相手チームの情報がないとき

- 「理由不十分の原則」に基づく事後確率

$$P(S_1|V) \approx 0.255$$
$$P(S_2|V) \approx 0.213$$
$$P(S_3|V) \approx 0.213$$
$$P(S_4|V) \approx 0.319$$

相手チームの情報があるとき

- 過去の戦績から、各戦術の選択率と勝率との関係は以下のとおりであった。



相手チームの勝率

- 各戦術を選択した条件下での勝率（尤度）と戦術選択率の合計＝勝率

$$P(V) = \sum_{i=1}^{N} P(V|Si)P(S_i)$$

$$= \frac{60}{100} \cdot \frac{100}{200} + \frac{5}{10} \cdot \frac{10}{200} + \frac{25}{50} \cdot \frac{50}{200} + \frac{30}{40} \cdot \frac{40}{200} = 0.6$$

相手チームの情報があるとき

- 戦術S_1を選ぶ事後確率

$$P(S_1|V) = \frac{P(V|S_1)P(S_1)}{P(V)} = \frac{(3/5)\cdot(1/2)}{0.6} = 0.5$$

相手チームの情報があるとき

- 各戦術を選ぶ事後確率

$$P(S_1|V) = 0.500$$
$$P(S_2|V) = 0.042$$
$$P(S_3|V) = 0.208$$
$$P(S_4|V) = 0.250$$

事後確率の比較

- 相手チームの情報がないとき

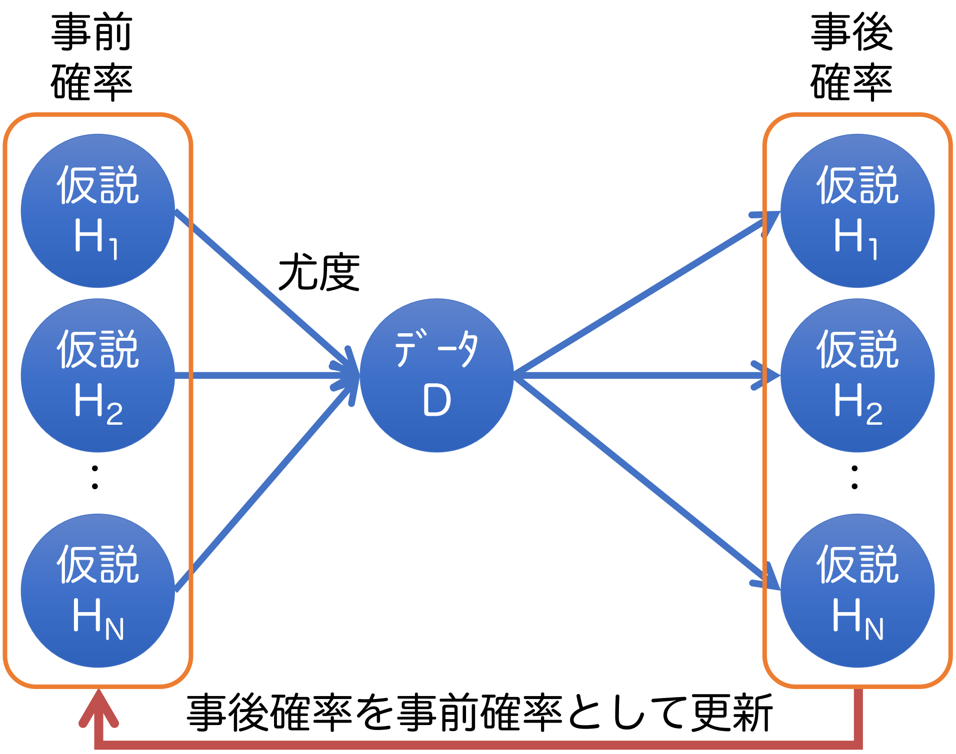
$$P(S_1|V) \approx 0.255$$

- 相手チームの情報があるとき

$$P(S_1|V) = 0.500$$

- 相手チームの情報の有無によって事後確率が異なるのはおかしくないか？

ベイズ更新



ベイズ更新

- 事後確率(0)＝尤度×事前確率(0)
- 事前確率(1)＝事後確率(0) ←更新1回目
- 事後確率(1)＝尤度×事前確率(1)
- 事前確率(2)＝事後確率(1) ←更新2回目
- 事後確率(2)＝尤度×事前確率(2)

:

- 事前確率(n)＝事後確率(n-1) ←更新n回目
- 事後確率(n)＝尤度×事前確率(n)

ベイズ更新の計算例

- 1 回目

更新回数

$$P^{(0)}(S_1) = P^{(0)}(S_1|V) \approx 0.255$$

$$P^{(1)}(S_1|V) = \frac{P^{(1)}(V|S_1)P^{(1)}(S_1)}{P^{(1)}(V)} \approx 0.253$$

$$P^{(1)}(S_2|V) \approx 0.176$$

$$P^{(1)}(S_3|V) \approx 0.176$$

$$P^{(1)}(S_4|V) \approx 0.395$$

ベイズ更新の計算例

- 2 回目

$$P^{(2)}(S_1) = P^{(1)}(S_1|V) \approx 0.253$$
$$P^{(2)}(S_1|V) = \frac{P^{(2)}(V|S_1)P^{(2)}(S_1)}{P^{(2)}(V)} \approx 0.243$$
$$P^{(2)}(S_2|V) \approx 0.141$$
$$P^{(2)}(S_3|V) \approx 0.141$$
$$P^{(2)}(S_4|V) \approx 0.475$$

クイズ

- ベイズ更新の 3 回目と 4 回目を計算してください。

ベイズ更新の計算例

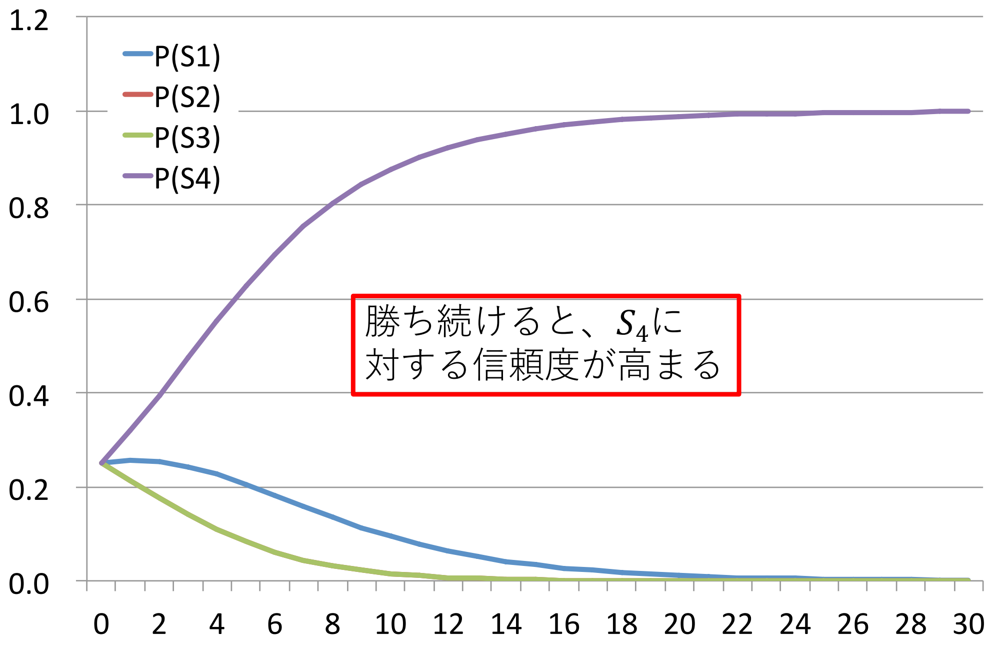
- 相手チームが勝ち続けていれば、戦術S_4が選ばれる確率が高くなる

尤度			
$P(V\|S_1)$	$P(V\|S_2)$	$P(V\|S_3)$	$P(V\|S_4)$
0.60	0.50	0.50	0.75

勝ち続けるとS_4への信頼度が高まる

更新回数	事前確率				勝率	事後確率			
	$P(S_1)$	$P(S_2)$	$P(S_3)$	$P(S_4)$	$P(V)$	$P(S_1\|V)$	$P(S_2\|V)$	$P(S_3\|V)$	$P(S_4\|V)$
0	0.250	0.250	0.250	0.250	0.588	0.255	0.213	0.213	0.319
1	0.255	0.213	0.213	0.319	0.605	0.253	0.176	0.176	0.395
2	0.253	0.176	0.176	0.395	0.624	0.243	0.141	0.141	0.475
3	0.243	0.141	0.141	0.475	0.643	0.227	0.109	0.109	0.554
:	:	:	:	:	:	:	:	:	:
29	0.002	0.000	0.000	0.998	0.750	0.001	0.000	0.000	0.999
30	0.001	0.000	0.000	0.999	0.750	0.001	0.000	0.000	0.999

ベイズ更新の計算例



ベイズ更新の計算例

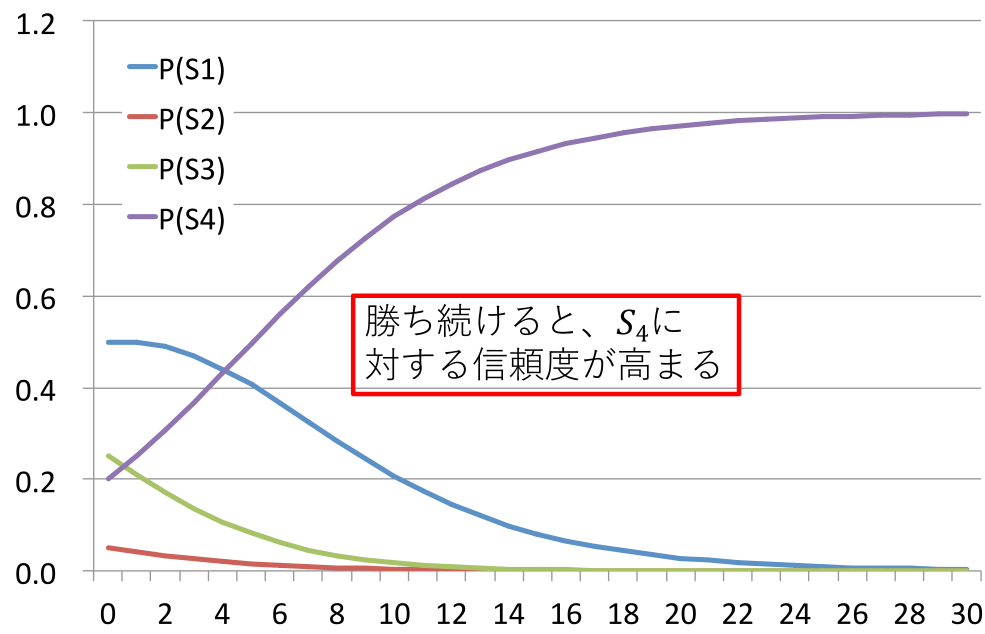
- 相手チームの情報があるとき
- 戦術S_4が選ばれる確率が高くなる

尤度			
$P(V\|S_1)$	$P(V\|S_2)$	$P(V\|S_3)$	$P(V\|S_4)$
0.60	0.50	0.50	0.75

勝ち続けると、S_4に対する信頼度が高まる

更新回数	事前確率				勝率	事後確率			
	$P(S_1)$	$P(S_2)$	$P(S_3)$	$P(S_4)$	$P(V)$	$P(S_1\|V)$	$P(S_2\|V)$	$P(S_3\|V)$	$P(S_4\|V)$
0	0.500	0.050	0.250	0.200	0.600	0.500	0.042	0.208	0.250
1	0.500	0.042	0.208	0.250	0.613	0.490	0.034	0.170	0.306
2	0.490	0.034	0.170	0.306	0.626	0.470	0.027	0.136	0.367
3	0.470	0.027	0.136	0.367	0.639	0.441	0.021	0.106	0.431
:	:	:	:	:	:	:	:	:	:
29	0.004	0.000	0.000	0.996	0.749	0.003	0.000	0.000	0.997
30	0.003	0.000	0.000	0.997	0.750	0.002	0.000	0.000	0.998

ベイズ更新の計算例



相手チームの戦術選択

ベイズ更新による計算の結果、

- 相手チームの情報がないとき
- 相手チームの情報があるとき

いずれの場合でも相手チームは戦術S_4を選び続けることで勝利Vに対する信頼度が高まり、戦術S_4を選択する確率が高くなることがわかった。

あなたのチームの戦術選択

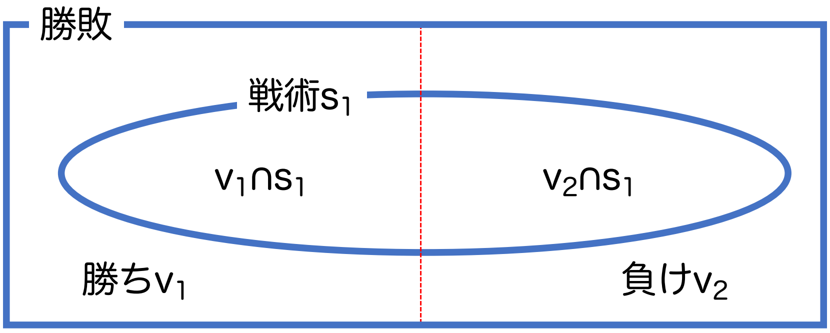
- 相手チームの戦術がわかったら、次はあなたのチームが戦術を考える番です。
- あなたのチームも4つの戦術(s_1, s_2, s_3, s_4)を使い分けることができる。
- アナリスト兼作戦立案者であるあなたは、ある作戦をとった条件下で試合に勝利する確率P(勝利|戦術)が最も高い作戦を選びたいと考えている。

あなたのチームの戦術選択

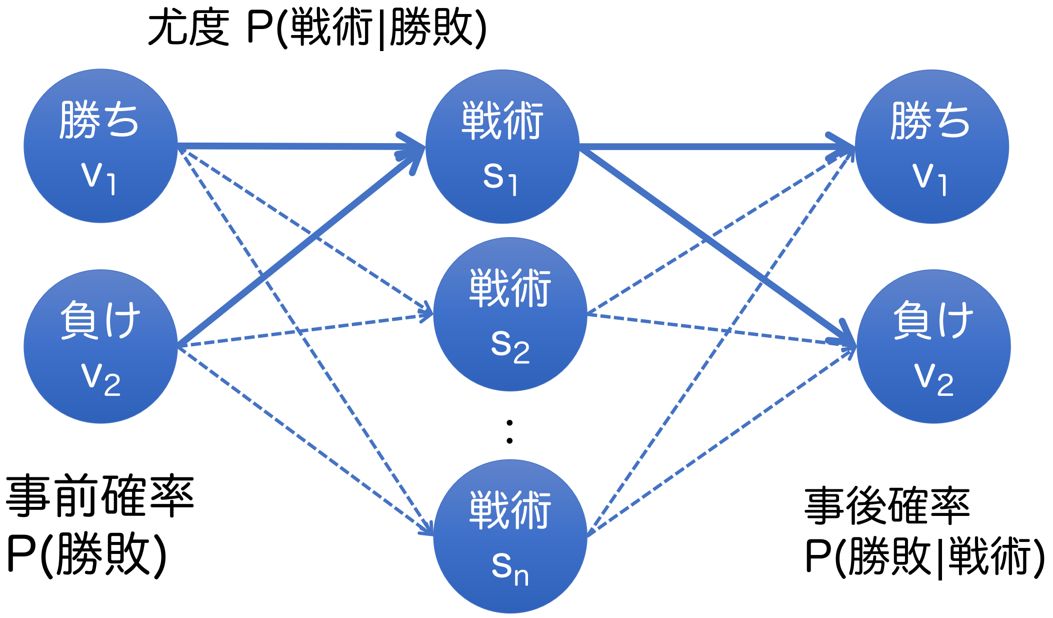
- 相手チームとは次の試合で初めて対戦することになるため、あなたのチームと相手チームの戦術の組み合わせによる勝敗に関するデータはありません。
- しかしこれまで他チームとの対戦経験で、相手チームが戦術S_4を選択したときに、あなたのチーム戦術に応じた勝敗数が分かっています。

あなたのチームの戦術選択

- あなたのチームが戦術s_1を選択した場合に試合に勝利(v_1)する事後確率$P(v_1|s_1)$



あなたのチームの戦術の検討



他チームとの対戦実績

- あなたのチームのこれまでの戦略別勝率は以下のとおりだとします。
- 全体としては勝ち試合数がやや上回っているものの、どの戦略も勝率は似たり寄ったりです。

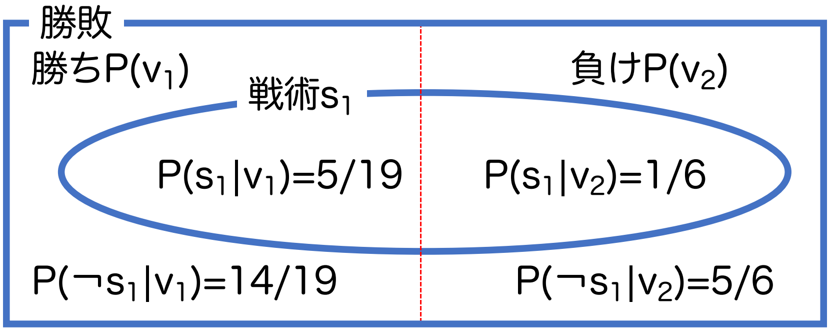
	s_1	s_2	s_3	s_4	合計
勝ちv_1	25	10	25	35	95
負けv_2	15	15	20	40	90
合計	40	25	45	75	185

あなたのチームの戦術選択

- 初めて対戦する相手とは勝敗の可能性はわからない。
- 他チームとは戦術を選んだ時の勝率はわかっている。
- これまでは戦術s_4を多く選んできたが、比較的勝率の高いs_1を選ぼうかと迷っている。
- 戦術s_4に対する信頼度は戦術s_1に対して高いといえるだろうか？

あなたのチームの戦術選択

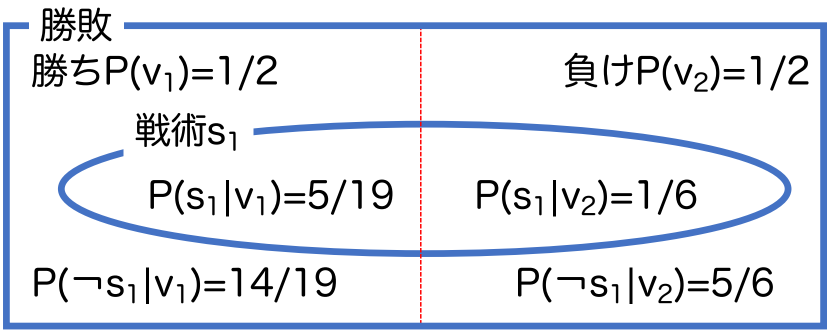
- 初めて対戦する相手とは勝敗の可能性はわからない。
- 他チームとは戦術s_1を選んだ時の勝率はわかっている。



対戦実績がないとき

- 初めて対戦する相手とは勝敗の可能性はわからない。
- 「理由不十分の原則」を適用する

$$P(v_1) = P(v_2) = 1/2$$



対戦実績がないとき

- 「理由不十分の原則」での事後確率は？

$$P(v_1|s_1) = \frac{P(s_1|v_1)P(v_1)}{P(s_1)}$$

$$= \frac{P(s_1|v_1)P(v_1)}{P(s_1|v_1)P(v_1) + P(s_1|v_2)P(v_2)}$$

$$= \frac{(5/19)\cdot(1/2)}{(5/19)\cdot(1/2) + (1/6)\cdot(1/2)} \approx 0.612$$

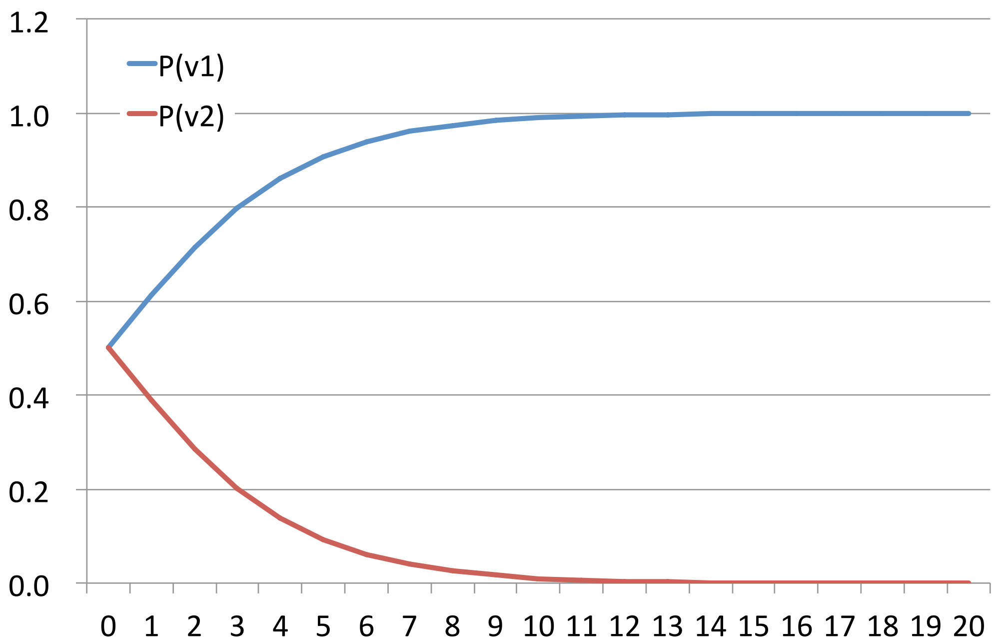
ベイズ更新の計算例

- 勝敗の実績がないとき
- 戦略s_1を選び続けると勝利に対する信頼度が高まる

尤度	
$P(s_1 \mid v_1)$	$P(s_1 \mid v_2)$
0.263	0.167

更新回数	事前確率		選択率	事後確率	
	$P(v_1)$	$P(v_2)$	$P(s_1)$	$P(v_1 \mid s_1)$	$P(v_2 \mid s_1)$
0	0.500	0.500	0.215	0.612	0.388
1	0.612	0.388	0.226	0.714	0.286
2	0.714	0.286	0.236	0.797	0.203
3	0.797	0.203	0.244	0.861	0.139
:	:	:	:	:	:
16	0.999	0.001	0.263	1.000	0.000
17	1.000	0.000	0.263	1.000	0.000

ベイズ更新の計算例



他チーム との対戦実績に基づくと

$$P(v_1|s_1) = \frac{P(s_1|v_1)P(v_1)}{P(s_1)}$$

$$= \frac{P(s_1|v_1)P(v_1)}{P(s_1|v_1)P(v_1) + P(s_1|v_2)P(v_2)}$$

$$= \frac{(5/19)\cdot(95/185)}{(5/19)\cdot(95/185) + (1/6)\cdot(90/185)} \approx 0.725$$

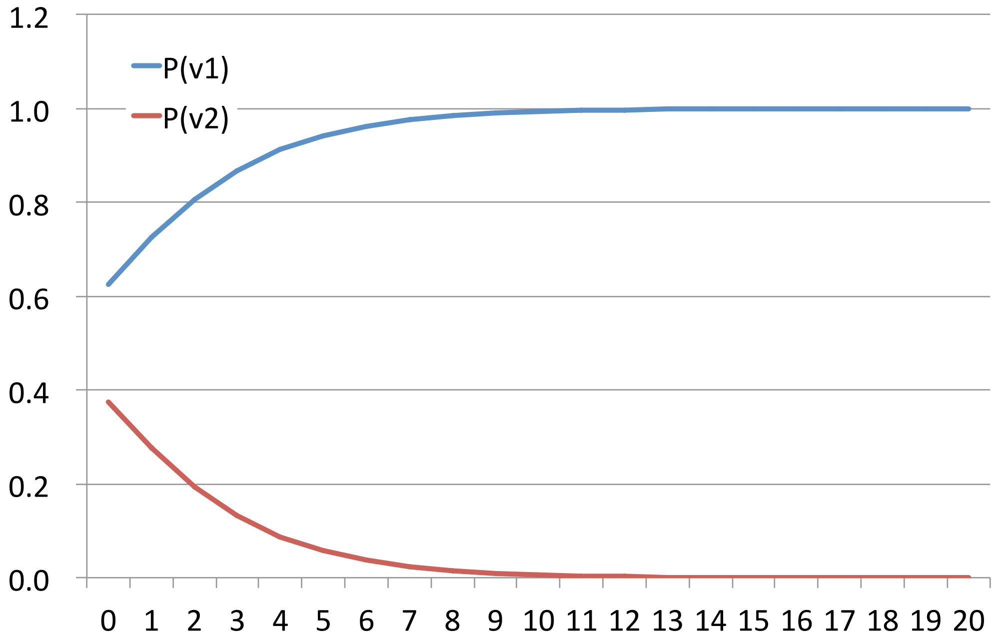
ベイズ更新の計算例

- 勝敗の実績があるとき
- 戦略s_1を選び続けると勝利に対する信頼度が高まる

尤度	
$P(s_1\|v_1)$	$P(s_1\|v_2)$
0.263	0.167

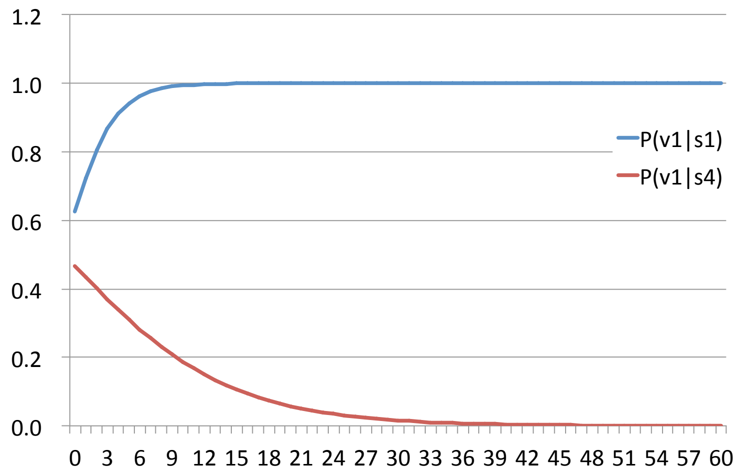
更新回数	事前確率		選択率	事後確率	
	$P(v_1)$	$P(v_2)$	$P(s_1)$	$P(v_1\|s_1)$	$P(v_2\|s_1)$
0	0.625	0.375	0.227	0.725	0.275
1	0.725	0.275	0.237	0.806	0.194
2	0.806	0.194	0.244	0.868	0.132
3	0.868	0.132	0.250	0.912	0.088
:	:	:	:	:	:
15	0.999	0.001	0.263	1.000	0.000
16	1.000	0.000	0.263	1.000	0.000

ベイズ更新の計算例



戦術同士の信頼度比較

- 戦術s_1に対して戦術s_4の信頼度は高いといえるだろうか？



実際に対戦すると…

- 実際の試合では、同じ戦術を選択しつづけるチームも、勝ち続けるチームもない
- 相手チームが戦術S_4を選択し続けるとして、あなたのチームは試行錯誤をすることとし、５試合で以下の戦術を取るとする。

$$s_1 \cdot \neg s_1 \cdot \neg s_1 \cdot s_1 \cdot s_1$$

- あなたは作戦立案者として、戦略s_1に対する勝利への確信度は薄らぐだろうか？

戦術s_1以外の尤度

- これまでは戦術s_1だけを選択し試合に勝つことだけを考えてきた
- 良い作戦立案者であるあなたは、
  - 試合に負ける場合
  - 他の戦術を選択した場合の勝敗
- についても検討するはず。

s_1の尤度		s_1以外の尤度	
$P(s_1 \| v_1)$	$P(s_1 \| v_2)$	$P(\neg s_1 \| v_1)$	$P(\neg s_1 \| v_2)$
0.263	0.167	0.737	0.833

5回の対戦結果によるベイズ更新

- 1回戦（＝更新回数0回）：戦術s_1

$$P^{(0)}(v_1|s_1) = \frac{P(s_1|v_1)P^{(0)}(v_1)}{P(s_1|v_1)P^{(0)}(v_1) + P(s_1|v_2)P^{(0)}(v_2)} \approx \underline{0.612}$$

$$P^{(0)}(v_2|s_1) = 1 - P^{(0)}(v_1|s_1) \approx \underline{0.388}$$

勝利することで戦術s_1に対する信頼度が高まった

5回の対戦結果によるベイズ更新

- 2回戦（＝更新回数1回）：$\neg s_1$

$$P^{(0)}(v_1|\neg s_1) = \frac{P(\neg s_1|v_1)P^{(0)}(v_1)}{P(\neg s_1|v_1)P^{(0)}(v_1) + P(\neg s_1|v_2)P^{(0)}(v_2)}$$
$$\approx \underline{0.583}$$

$$P^{(0)}(v_2|\neg s_1) = 1 - P^{(0)}(v_1|\neg s_1) \approx \underline{0.417}$$

勝率が下がることで戦術s_1以外の選択に対する信頼度が下がる

5 回の対戦によるベイズ更新

- 戦術を変更することで勝率（勝利に対する信頼）が下がる。戦術を戻すと勝率（勝利に対する信頼）が上がる。

試合回数	更新回数	戦術	事前確率		選択率	事後確率	
			$P(v_1)$	$P(v_2)$	$P(s_1)$	$P(v_1\|s_1)$	$P(v_2\|s_1)$
1	0	s_1	0.500	0.500	0.215	0.612	0.388
2	1	$\neg s_1$	0.612	0.388	0.774	0.583	0.417
3	2	$\neg s_1$	0.583	0.417	0.777	0.552	0.448
4	3	s_1	0.552	0.448	0.220	0.661	0.339
5	4	s_1	0.661	0.339	0.230	0.755	0.245

ベイズ更新からベイズ統計へ

- ベイズ更新は迷惑メールのフィルタリングなどに用いられる

- 変数（データ）が多く存在するとき、事前確率や尤度は離散的な確率分布だけでなく、連続的な確率分布になる

- 統計データのモデル推定にベイズ更新を応用したものがベイズ統計である