

# 裏切られる直感と、 確率と精度のはなし

---

K. Yamaguchi

2019-01-21

- ✓ モンティ・ホール問題
- ✓ ウィルス感染検査装置の問題
- ✓ システムの評価尺度のはなし

# モンティ・ホール問題

A



B



C



## テレビ番組の景品当てゲーム

東京フレンドパークのダーツとか

アタック25の海外旅行チャレンジとか

# 扉の向こうには.....

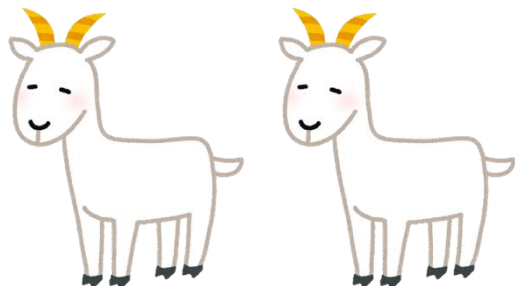
A



B



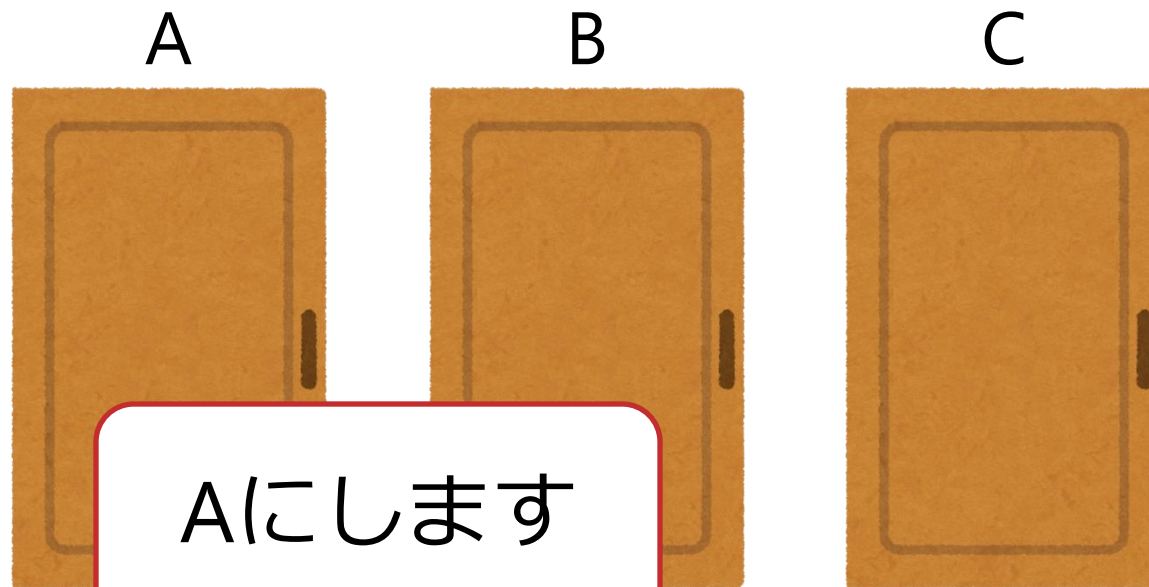
C



当たりが1つ

はずれが2つ

# 回答者が1つを選ぶ



Aにします



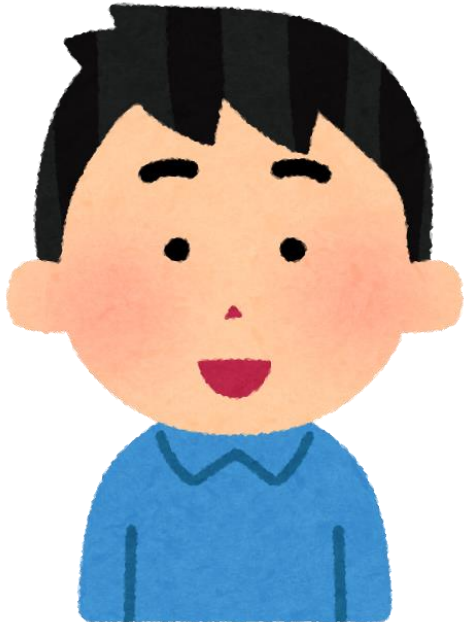
A



B



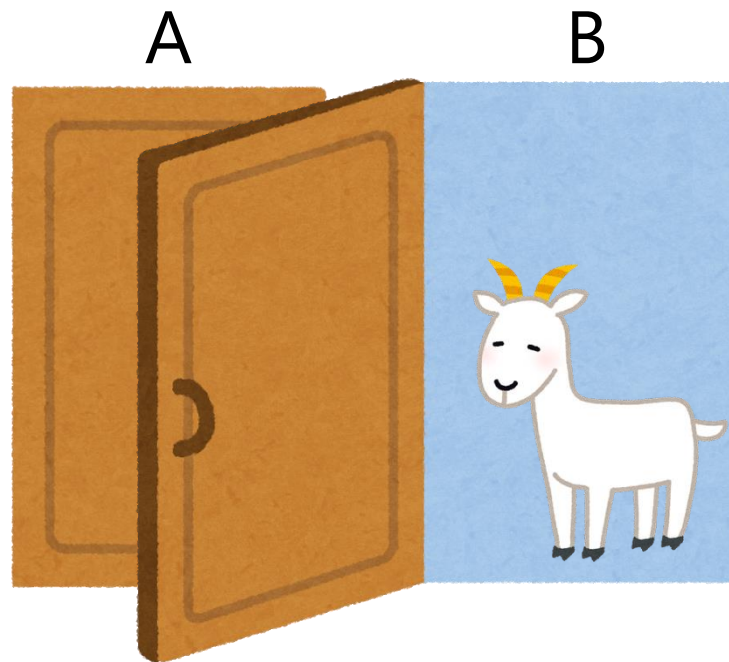
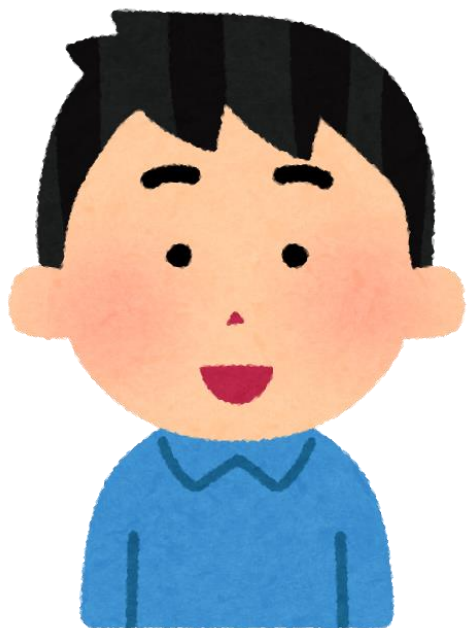
C



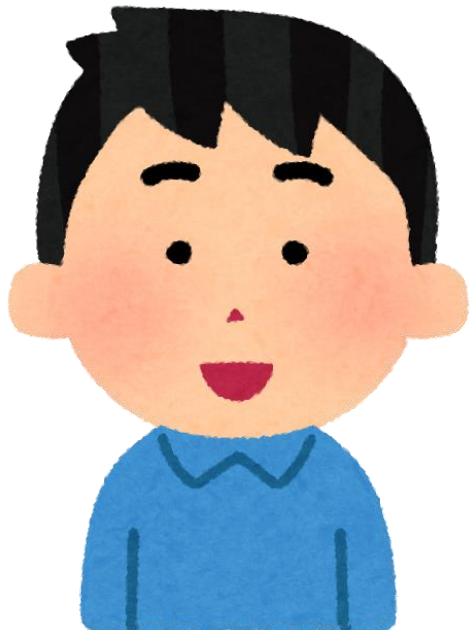
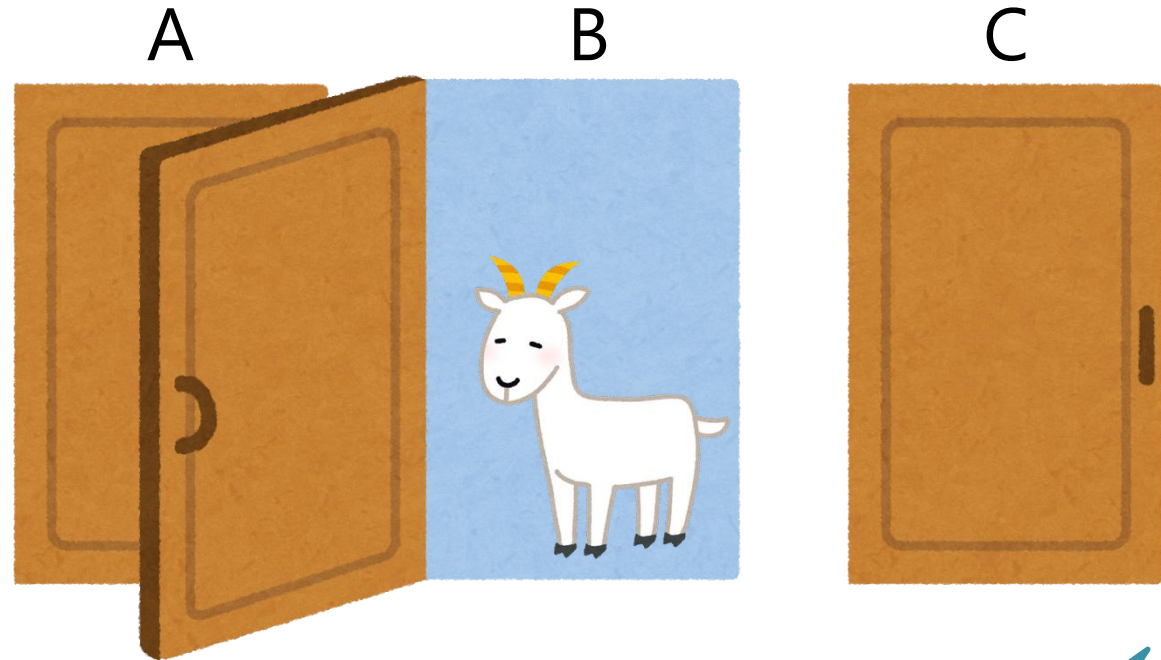
A以外の2つのうち  
**1つを開けて**  
見せてあげましょう



# 司会者が扉を開け1つのはずれが判明する



# 選び直すことができる」と司会者は言う



あなたはもう1回  
別の扉を選び直す  
ことができますよ



はたして

Aのままにするべきだろうか？

それとも

Cに選び直すべきだろうか？

荷々々々

まったくよくない出題です

1990年9月9日発行、ニュース雑誌「Parade」にてマリリン・ボス・サヴァントが連載するコラム「マリリンにおまかせ」で、上記の読者投稿による質問に ここはのちほど と回答した。すると直後から、読者からの「彼女の解答は間違っている」との約1万通の投書が殺到し、本問題は大きな議論に発展した。

(略)

投書には、1000人近い博士号保持者からのものも含まれていた。

人間の直感を裏切る問題

だけでなく

致命的な部分で出題があいまい

という面も

1. 回答者には3つの選択肢が与えられており、そのうちの1つが当たり、2つが外れである
2. 司会者がいて当たりがどれかを事前に知っている
3. 回答者は選択肢から1つを選び司会者に伝える
4. 司会者は状況によって以下のどちらかの行動を取る
  1. 回答者が外れを選択した場合、残った2つの選択肢のうち外れの方を取り除く
  2. 回答者が当たりを選択した場合、残った2つの選択肢のうち一方を無作為に（つまりそれぞれ等しい確率で）取り除く
5. 回答者は最初に選んだ選択肢と、取り除かれずに残った選択肢からもういちど選び直すことができる
6. 上記のルール・条件は事前に回答者に知らされている

正解は『選択肢を変更する』である

なぜなら選択肢を変更した場合には  
景品を当てる確率が2倍になるから

# どう考えるとわかりやすいか？

このゲームを900人ぐらいがチャレンジする

Aを選択

300人ぐらい

Bを選択

300人ぐらい

Cを選択

300人ぐらい

# どう考えるとわかりやすいか？

このゲームを900人ぐらいがチャレンジする

300人ぐらいがAを選択する

Aを選択

300人ぐらい

# どう考えるとわかりやすいか？

このゲームを900人ぐらいがチャレンジする

300人ぐらいがAを選択する

Aが正解  
100人ぐらい

Bが正解  
100人ぐらい

Cが正解  
100人ぐらい

# どう考えるとわかりやすいか？

このゲームを900人ぐらいがチャレンジする

300人ぐらいがAを選択する

Aが正解で <u>Bが除外</u> 50人ぐらい	Aが正解で <u>Cが除外</u> 50人ぐらい
Bが正解 → <u>Cが除外</u> 100人ぐらい	
Cが正解 → <u>Bが除外</u> 100人ぐらい	

# どう考えるとわかりやすいか？

このゲームを900人ぐらいがチャレンジする

300人ぐらいがAを選択する

150人ぐらいが、Aを選択してBを除外される状況を観測する

Aが正解で  
Bが除外  
50人ぐらい

Cが正解 → Bが除外  
100人ぐらい

# どう考えるとわかりやすいか？

- このゲームに900人がチャレンジすると自分と同じ状況になる人が150人いて
- その中の50人は選択肢を変えない時に当たる
- その中の100人が選択肢を変える時に当たる

選択肢を変えると当たる人が2倍

1. 回答者には3つの選択肢が与えられており、そのうちの1つが当たり、2つが外れである
2. 司会者がいて当たりがどれかを事前に知っている
3. 回答者は選択肢から1つを選び司会者に伝える
4. 司会者は状況によって以下のどちらかの行動を取る
  1. 回答者が外れを選択した場合、残った2つの選択肢のうち外れの方を取り除く
  2. 回答者が当たりを選択した場合、残った2つの選択肢のうち一方を無作為に（つまりそれぞれ等しい確率で）取り除く
5. 回答者は最初に選んだ選択肢と、取り除かれずに残った選択肢からもういちど選び直すことができる
6. 上記のルール・条件は事前に回答者に知らされている

# ウイルス感染検査装置の問題

- あるウィルスに感染しているかどうかを検査できる装置がある
- 装置が判定した結果の正解率は99%
- 1,000人に1人の割合で潜在的感染者がいる
- あなたがその検査装置を使ったところ陽性という判定が出た

この条件で  
**本当にウィルスに感染している**  
確率はどれくらい?

9%ぐらい

# どう考えるとわかりやすいか？

この装置で10万人ぐらいを検査する

100人ぐらいが感染者

感染者100人

非感染者99,900人



比率は正しくない

# どう考えるとわかりやすいか？

この装置で10万人ぐらいを検査する

100人ぐらいが感染者

感染者100人

非感染者99,900人

感染者  
かつ  
結果が陽性  
99人

非感染者  
かつ  
結果が陽性  
999人

比率は正しくない

99人の感染者かつ結果が陽性

999人の非感染者かつ結果が陽性

真に感染している確率 =

$$\frac{\text{真に感染}}{\text{結果が陽性}} = \frac{99}{99 + 999} = \frac{99}{1,098} \doteq 0.09$$

9%

おきづきだった  
だろうか……？

常に「感染していない」  
と

答えていれば正解率が

**99.9%**

だということに……

# 正解率はあてにならない

- とはいえさすがに常に「感染していない」と答えるのではシステムとして評価できない
- しかし

非感染者の割合99.9%で  
ランダムに感染していない  
と答える

感染者の割合0.01%で  
ランダムに感染している  
と答える

## ようなシステム

- 正解率は99.8%ある
  - $0.999 \times 0.999 + 0.001 \times 0.001 = 0.998$

**マジョリティ  
ベースライン**  
という名前が  
ついている

# 正解率はあてにならない

Aであるかを判定する問題で、Aである/ないの割合が

**極端に違う時**

正解率という指標は**当てにならない**

## □ 例えば

- 異常検知
- 迷惑メール判定
- 群衆の中での特定人物の識別

判定において正解率約8割の  
精度を確認しました

精度とは?



システムの評価尺度のはなし

## 適合率

(precision)

Aであると判定したもののうち  
本当にAである割合

## 再現率

(recall)

本当にAであるもののうち  
Aであると判定できた割合

※現実の問題では  
真の再現率はわからない  
ことの方が多い

## 適合率

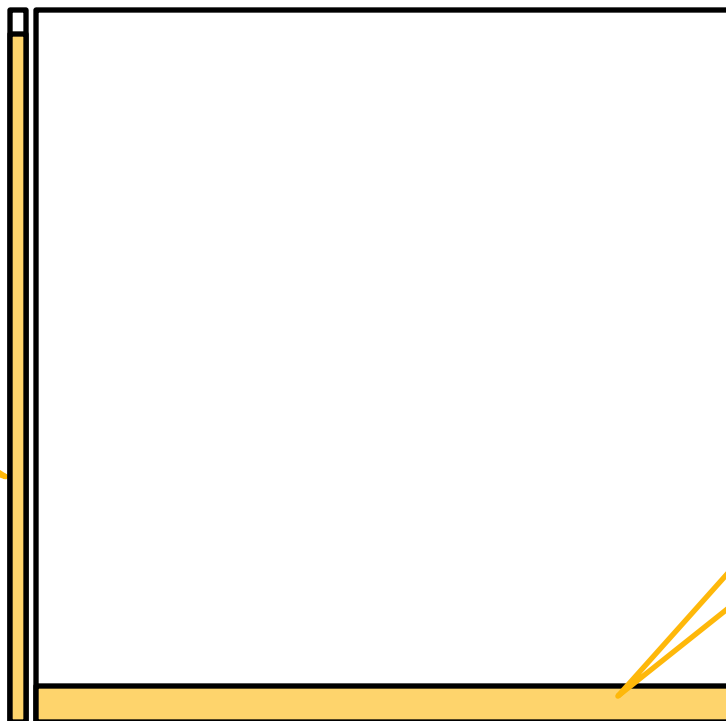
(precision)

Aであると判定したもののうち  
本当にAである割合

$$\frac{99}{99 + 999} \doteq 0.09$$

感染者  
かつ  
結果が感染  
99人

非感染者  
かつ  
結果が感染  
999人



## 再現率

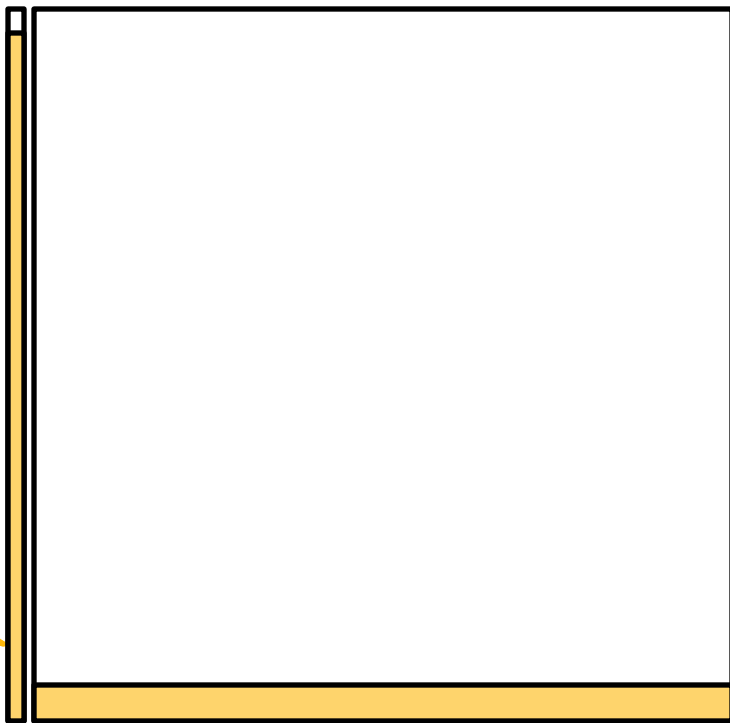
(recall)

本当にAであるもののうち  
Aであると判定できた割合

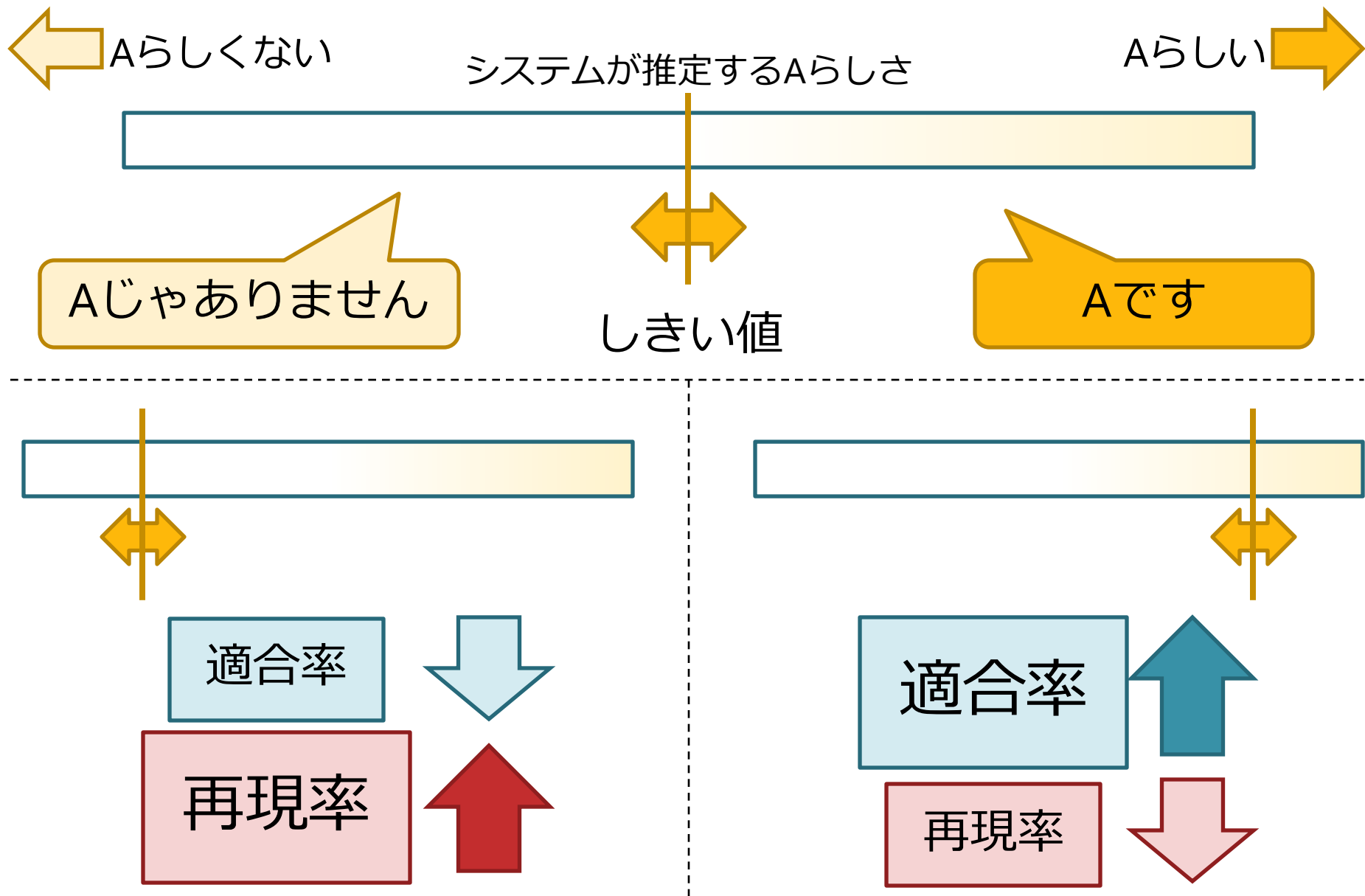
$$\frac{99}{100} = 0.99$$

感染者 100人

感染者  
かつ  
結果が感染  
99人



# 適合率と再現率はトレードオフ



適合率と再現率の調和平均という尺度

$$\frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

どちらかが極端に下がると値が下がる

||

いい案配にバランスが取れていると値が上がる

## マジョリティベースライン

正解率 : 0.998

適合率 : 0.001  
(precision)

再現率 : 0.001  
(recall)

F1値 : 0.0001

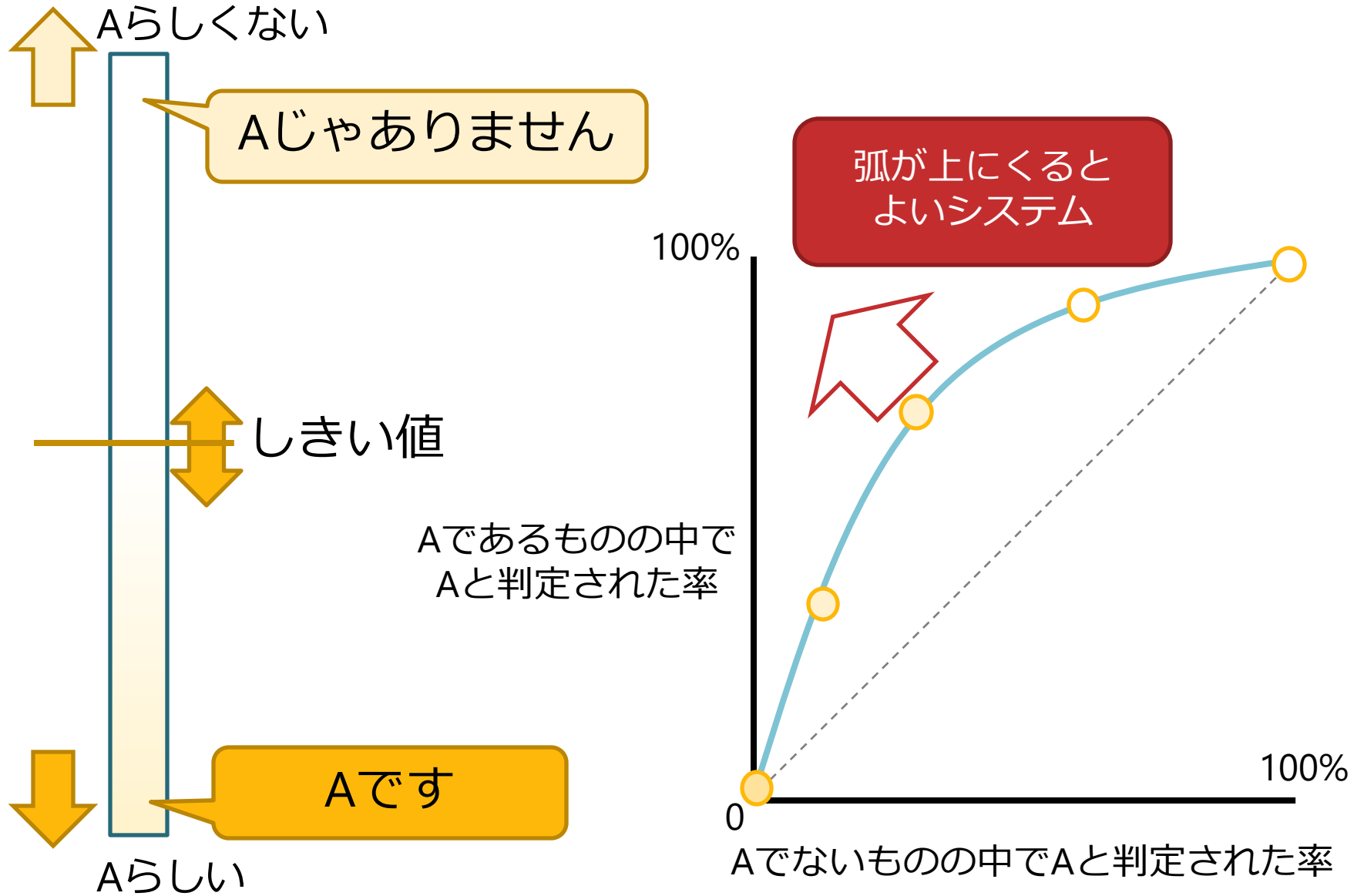
## 正解率99%のシステム

正解率 : 0.99

適合率 : 0.09  
(precision)

再現率 : 0.99  
(recall)

F1値 : 0.165



判定において正解率約8割の  
精度を確認しました

F1値?

適合率?

正解率?



まとめ

- 条件付き確率の問題は容易に直感を裏切る
- Aである/ないに極端に差がある時は特に注意  
正解率はよい指標にならない
- 適合率と再現率という指標がある  
どちらを重視するかはシステムのさじ加減
- F値というバランスのよさを測る指標がある