

統計学 2

2. Rを使ってみよう

矢内 勇生

2019年4月15日

高知工科大学 経済・マネジメント学群

今日の目標

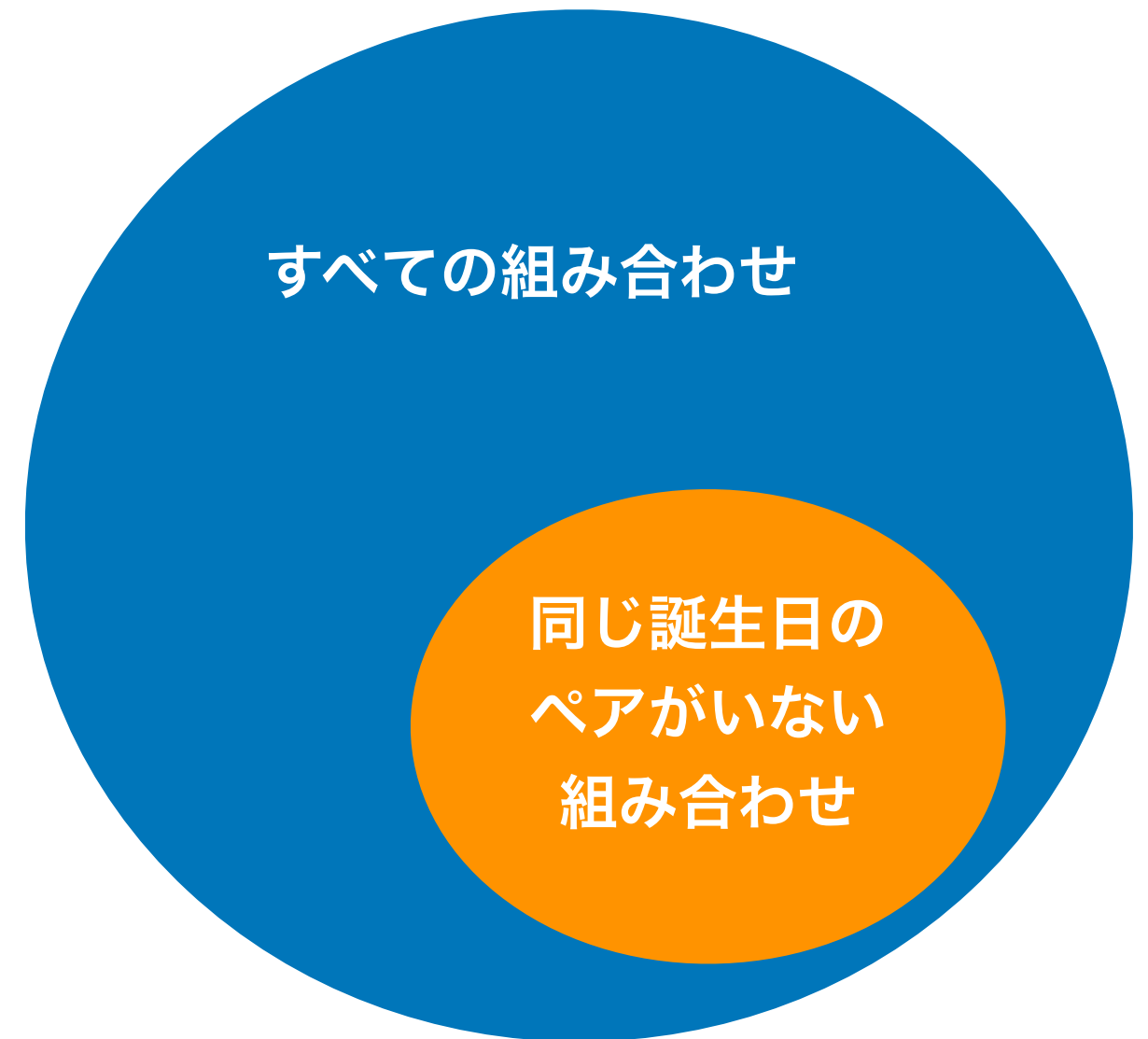
- Rを使ってみる！
 - ▶ Rが便利な道具であることを知る
 - ▶ Rの難しさについて知る

誕生日問題

- 40人のクラスに誕生日が同じ人がいる確率は？
 - － 直感：誕生日は365通りあり、40人しかいないから確率は低いはず。
 - － 事実：約90%

少なくとも1組同じ誕生日がいる確率

- すべての確率 (1 = 100%) から同じ誕生日のペアがない確率を引けばよい



求めたい確率 = 1 - 同じ誕生日のペアが1組もない確率

1組も同じ誕生日のペアがない場合は何通り？

- 40人のクラスで誕生日が重複しないような組み合わせ
 - ▶ 年は考えない
 - ▶ 2月29日は除外
- 1人目の選び方：365通り（どの誕生日でもよい）
- 2人目の選び方：まだ選ばれていない364日から選ぶ
- 3人目の選び方：まだ選ばれていない363日から選ぶ
- . . .
- 40人目の選び方：まだ選ばれていない326日から選ぶ

1組も同じ誕生日のペアがない場合の数

$$365 \times 364 \times 363 \times \cdots \times 326$$

40コの数进行かける！

誕生日の組み合わせは全部で何通り？

- それぞれの人について、365通りから選べる

$$365^{40}$$

365を40乗する（40回かける）！

40人の中に1組も同じ誕生日のペアがない確率

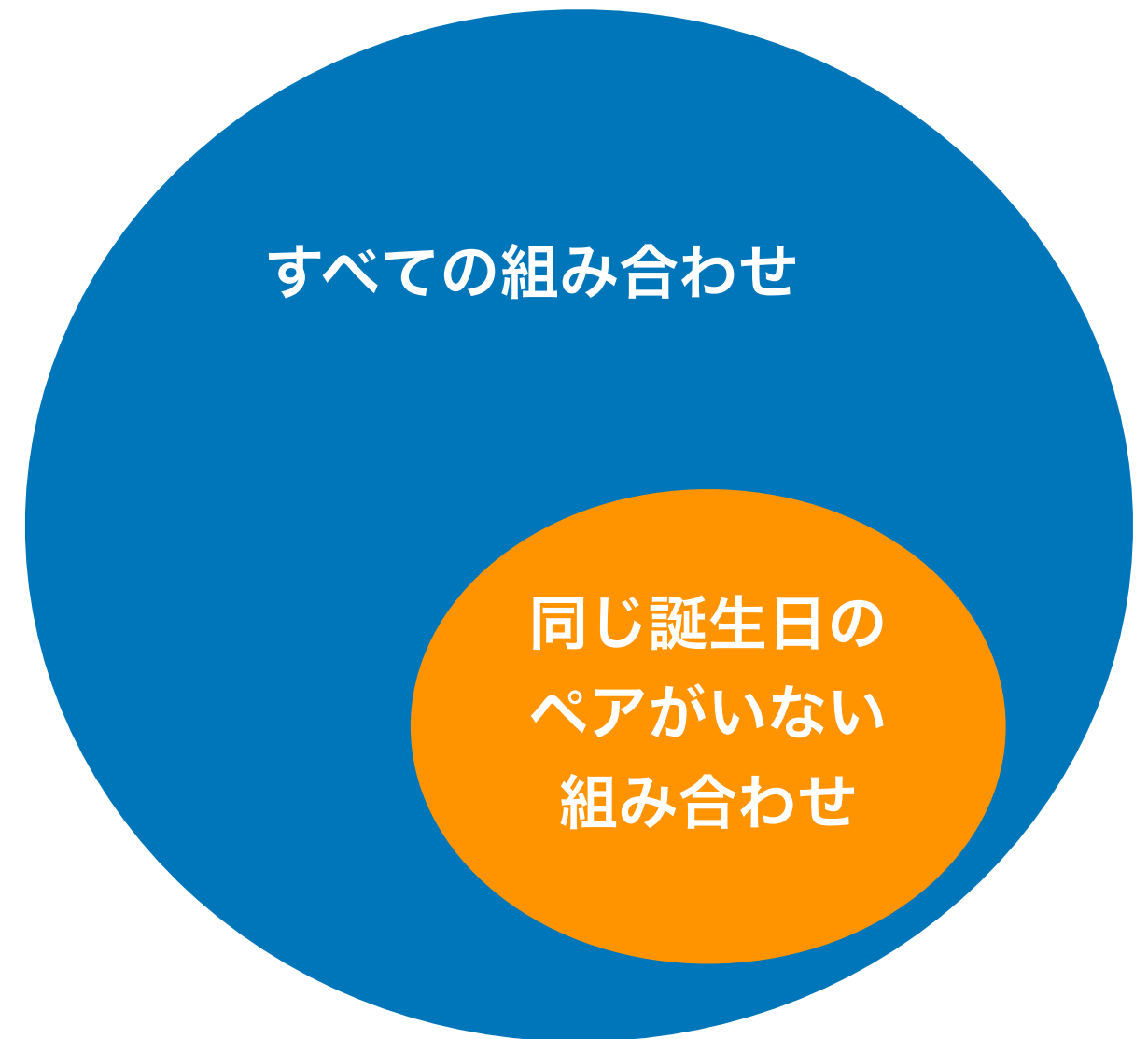
1組も同じ誕生日がない組み合わせ

すべての組み合わせ

$$= \frac{365 \times 364 \times \cdots \times 326}{365^{40}}$$

少なくとも1組同じ誕生日がいる確率

- すべての確率 (1 = 100%) から同じ誕生日のペアがない確率を引けばよい



$$\text{求めたい確率} = 1 - \frac{365 \times 364 \times \cdots \times 326}{365^{40}}$$

どうやって計算する？

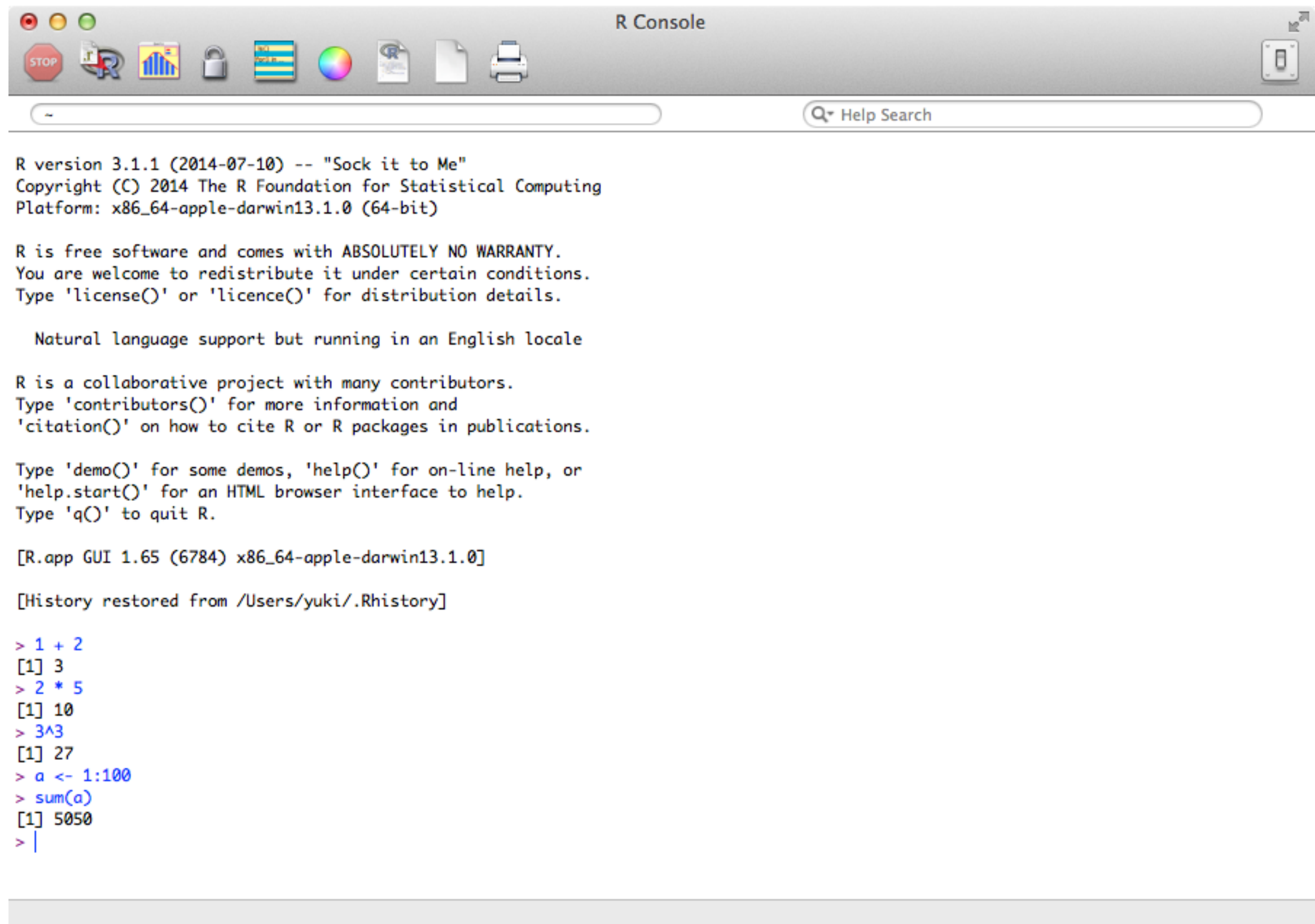
- 暗算する
- 手で計算（筆算）する
- 課題：コンピュータを使って答えを求めよう！
 - ▶ 現時点での自分のスキルを使って挑戦
 - ▶ R (RStudio) で計算する
 - ▶ R (RStudio) でもっと楽に計算する：関数の利用



とは何か？

- この授業でメインで使うソフトウェア
- 統計学の研究においては標準ソフト
- オープンソースでオブジェクト指向プログラミング言語
- Comprehensive R Archive Network (CRAN) で入手可能

Rのスクリーンショット



```
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.1.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.65 (6784) x86_64-apple-darwin13.1.0]

[History restored from /Users/yuki/.Rhistory]

> 1 + 2
[1] 3
> 2 * 5
[1] 10
> 3^3
[1] 27
> a <- 1:100
> sum(a)
[1] 5050
> |
```

Rのメリット

- 無料
 - Stata (2万円 [学生版], 12万円 [一般]), SPSS (10万円 [教育], ??? [一般]), etc.
- 柔軟：ほとんどなんでもできる
 - 追加のパッケージは無料でインストール可能
- 世界中で様々な分野の研究者に広く利用されている
 - 共同研究の促進, 研究の相互チェック, etc.
- 美しいグラフが描ける
 - 研究成果を効果的に伝えるために重要

Rのデメリット

- 設定が面倒
- プログラミングが必要
 - 少しでも間違えると、動かない or 意図せざる動作
 - 英語を使う必要がある（メリットにもなり得る）
- ★一言でいうと「難しい」（Rのデメリットというよりは、プログラミング一般のデメリット）

R導入の手引き書

- Rの導入について詳しく解説
- 自分のパソコンにRを導入したい人にオススメ



モンティ・ホール問題

テレビのクイズ番組

- 3つのドアのうち1つのドアの後ろにランダムに当たり（景品）を置く
- あなたは好きなドアを選ぶ
- あなたが1つ選んだ後、司会者がはずれを1つ教えてくれる
- あなたは最初に選んだドアにとどまりますか？ドアを変えますか？

3つの可能性

- 変えないほうがよい
 - 変えたほうがよい
 - どちらも同じ
- ◆ どれが正しい？

ドアは変えるべき！

- 当たりの確率が2倍になる！

解法1

- 最初に当たりを引く確率は $1/3$, はずれを引く確率は $2/3$ (はずれを区別するとわかりやすい: 例) たわしとスポンジ)
 - 最初が当たり: とどまれば確実に (確率1で) 当たり、移動すると確実にはずれる (確率0で当たる)
 - 最初がはずれ: とどまれば確実にはずれ (確率0である)、移動すれば確実に (確率1で) 当たり
- ➡ とどまれば $1/3$ 、移動すれば $2/3$ である

解法2

- 自分が選んだドアが当たりの確率は $1/3$ 、それ以外が当たりの確率は $2/3$
- 自分が選んでいないドアのどちらかが開けられても、自分が選んだドア以外が当たる確率は $2/3$ のまま
- とどまれば $1/3$ 、移動すれば $2/3$ で当たる

誕生日問題とモンティ・ホール問題の教訓

- 直感で正しいと思う解答と、論理的に正しい解答は異なる

★ 直感は間違える！

- ➡ 確率論や統計学など、直感に頼らない方法を身につけることが必要