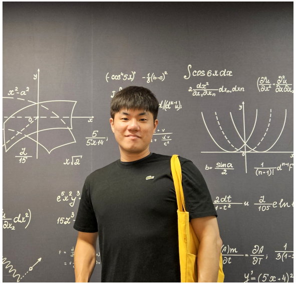
# Probabilistic Modelling: Ruin Probability and the Monty Hall Problem

국가수리과학연구소 김민중 (Ph.D.)

Probability, Random Walks, and Conditional Probability

Stochastic Processes I Risk Modelling I Information and Probability

# About me



**김민중 (Ph. D.)**

국가수리과학연구소(**NIMS**) 산업수학혁신센터(**ICIM**) 선임연구원 **김민중**입니다. 확률론을 전공하였으며, 수학과 산업의 접점에서 데이터 분석, 머신러닝, 딥러닝 기반 협력 연구를 수행하고 있습니다.

최근에는 **수학 기반 인공지능** 연구에 집중하며, 다음 주제를 중심으로 연구하고 있습니다.

## 주요 연구 분야

- **생성 모델 개선**: 생성 모델의 구조를 수학적으로 분석하고 알고리즘을 개선
- **확률해석 기반 SciML**: SOC와 DRO 문제를 PDE로 정식화하고, 과학기계학습을 활용한 해 검증 연구

## Work Experience

- **국가수리과학연구소** 2016 –
- **아주대학교 겸임교수** 2022 –

# Outline

- Part I. Ruin probability modelling
  - a casino story: from 10 dollars to 80 dollars
  - random walk model and ruin probability
  - recursive equation and explicit formula
  - constant bet size and state-dependent strategies
  - bold play and strategy comparison
  - interpretation in insurance mathematics
- Part II. The Monty Hall problem
  - why switching is better
  - conditional probability
  - 4-door and $N$-door generalizations

## A simple story

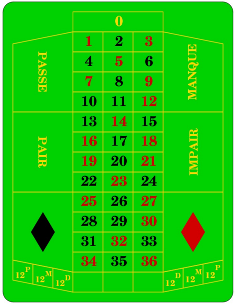
Suppose that you immediately need $\$80$, but at the moment you only have $\$10$.

- You decide to go to a casino.
- At the roulette table, you repeatedly bet on **red**.
- Each time, you bet $\$1$.
- You stop as soon as one of the following happens:
  - you reach $\$80$,
  - or you lose everything.

The main question is:

What is the probability that you lose all your money before reaching $\$80$?

# Why Roulette?




- In European roulette, there are $37$ slots:
  - $18$ red,
  - $18$ black,
  - $1$ green zero.
- So if you bet on red, then

$$p = P(\text{win}) = \frac{18}{37}, \qquad q = P(\text{lose}) = \frac{19}{37}.$$

- Therefore the game is slightly unfavorable:

$$\frac{18}{37} < \frac{1}{2}.$$

- This small disadvantage becomes important when the game is repeated many times.

# From the Story to a Mathematical Model

- Let $X_n$ be your capital after the $n$-th game.
- Initially,

$$X_0 = 10.$$

- Each game changes your capital by $\pm 1$:

$$X_{n+1} = \begin{cases} X_n + 1, & \text{with probability } p = \frac{18}{37}, \\ X_n - 1, & \text{with probability } q = \frac{19}{37}. \end{cases}$$

- The game stops when the process hits either

$$0 \qquad \text{or} \qquad 80.$$

This is a simple random walk with absorbing boundaries.

# Ruin Probability

- For a general initial capital $i$, define

$$R(i) = P(\text{hit } 0 \text{ before } 80 \mid X_0 = i).$$

- This is called the **ruin probability**.
- In our original story, the quantity of interest is

$$R(10).$$

- So the casino question becomes:

$$\text{Starting from } \$10, \text{ what is the probability of ruin before reaching } \$80?$$

# Boundary Conditions and Recursion

- If you already have $\$0$, then ruin has already happened:

$$R(0) = 1.$$

- If you already have $\$80$, then ruin is impossible:

$$R(80) = 0.$$

- For $1 \leq i \leq 79$, conditioning on the first game gives

$$R(i) = pR(i+1) + qR(i-1),$$

  where $p = \frac{18}{37}$ and $q = \frac{19}{37}$.

- Thus ruin probability satisfies a discrete boundary value problem.

# Solving the Recursive Equation: Step 1

For $1 \leq i \leq 79$, the ruin probability satisfies

$$R(i) = pR(i+1) + qR(i-1).$$

Since $p + q = 1$, we also have

$$R(i) = pR(i) + qR(i).$$

Subtracting the recursion from this identity gives

$$p\{R(i) - R(i+1)\} = q\{R(i-1) - R(i)\}.$$

Therefore,

$$R(i) - R(i+1) = \frac{q}{p}\{R(i-1) - R(i)\}.$$

This shows that the successive differences form a geometric sequence.

# Solving the Recursive Equation: Step 2

In the roulette example,

$$\frac{q}{p} = \frac{19/37}{18/37} = \frac{19}{18}.$$

Thus

$$R(i) - R(i+1) = \frac{19}{18}\{R(i-1) - R(i)\}.$$

For example,

$$R(1) - R(2) = \frac{19}{18}\{R(0) - R(1)\},$$

$$R(2) - R(3) = \left(\frac{19}{18}\right)^2 \{R(0) - R(1)\}.$$

In general,

$$R(i) - R(i+1) = \left(\frac{19}{18}\right)^i \{R(0) - R(1)\}.$$

# Solving the Recursive Equation: Step 3

Now use the telescoping sum:

$$R(0) - R(80) = \sum_{i=0}^{79} \{R(i) - R(i+1)\}.$$

Since $R(0) = 1$ and $R(80) = 0$,

$$1 = \sum_{i=0}^{79} \left(\frac{19}{18}\right)^{i} \{R(0) - R(1)\}.$$

Using the geometric sum,

$$1 = \frac{1 - \left(\frac{19}{18}\right)^{80}}{1 - \frac{19}{18}} \{1 - R(1)\}.$$

This determines $R(1)$, and then all other $R(i)$.

# Explicit Formula for the 1-Dollar Betting Strategy

For the roulette example with $K = 80$,

$$R(i) = \frac{\left(\frac{19}{18}\right)^i - \left(\frac{19}{18}\right)^{80}}{1 - \left(\frac{19}{18}\right)^{80}}, \qquad i = 0, 1, \ldots, 80.$$

In particular, starting from $\$10$,

$$R(10) = \frac{\left(\frac{19}{18}\right)^{10} - \left(\frac{19}{18}\right)^{80}}{1 - \left(\frac{19}{18}\right)^{80}} \approx 0.9904.$$

Therefore,

$$1 - R(10) \approx 0.0096.$$

So betting $\$1$ each time gives less than $1\%$ chance of success.

# General Formula

- More generally, if the target is $K$, the win probability is $p$, and $q = 1 - p$, then

$$R(i) = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^K}{1 - \left(\frac{q}{p}\right)^K}, \qquad p \neq \frac{1}{2}.$$

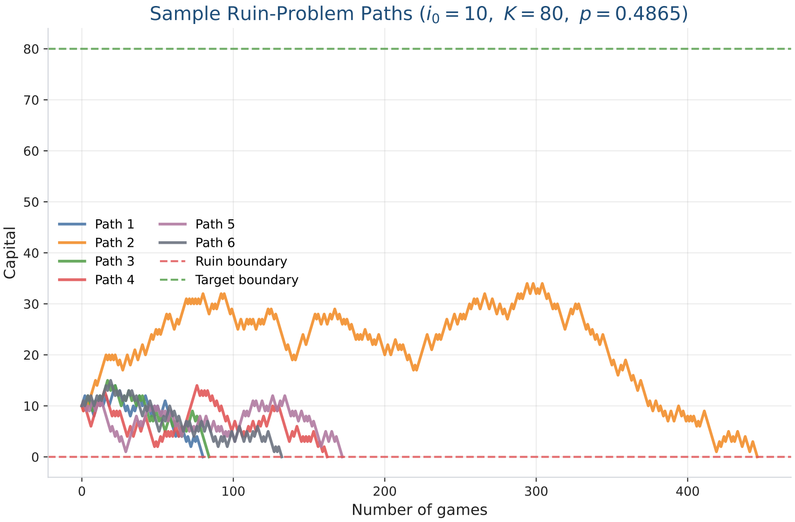
- In the fair case $p = \frac{1}{2}$,

$$R(i) = 1 - \frac{i}{K}.$$

- This formula comes from the same recursive equation and boundary conditions:

$$R(0) = 1, \qquad R(K) = 0.$$

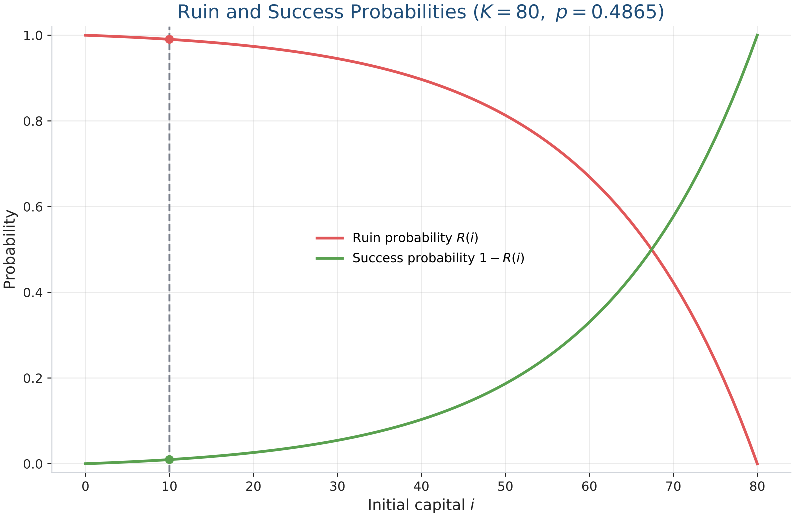
# Sample Paths of the Capital Process



# Interpretation of Sample Paths

- Each path begins at $\$10$.
- A win moves the capital up by $1$, and a loss moves it down by $1$.
- The lower horizontal line is the ruin boundary $0$.
- The upper horizontal line is the target boundary $80$.
- Since the roulette game is unfavorable, many paths fall to $0$ before ever reaching $80$.

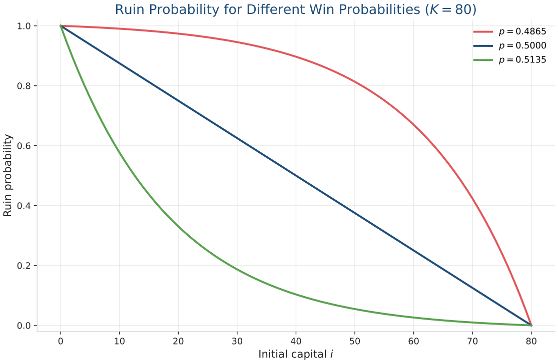
# Ruin and Success Probabilities

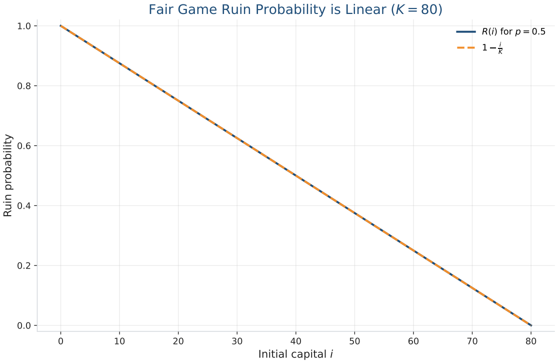


# Reading the Probability Curve

- The graph of $R(i)$ shows the probability of losing everything before reaching $80$.
- The graph of $1 - R(i)$ shows the probability of success.
- The marked point $i = 10$ corresponds to our original casino story.
- Since $p = \frac{18}{37} < \frac{1}{2}$, the ruin probability remains very high unless the initial capital is already large.

# Unfair, Fair, and Favorable Games




Different values of $p$ lead to different ruin profiles.

# Expected Drift

- The expected change in one step is

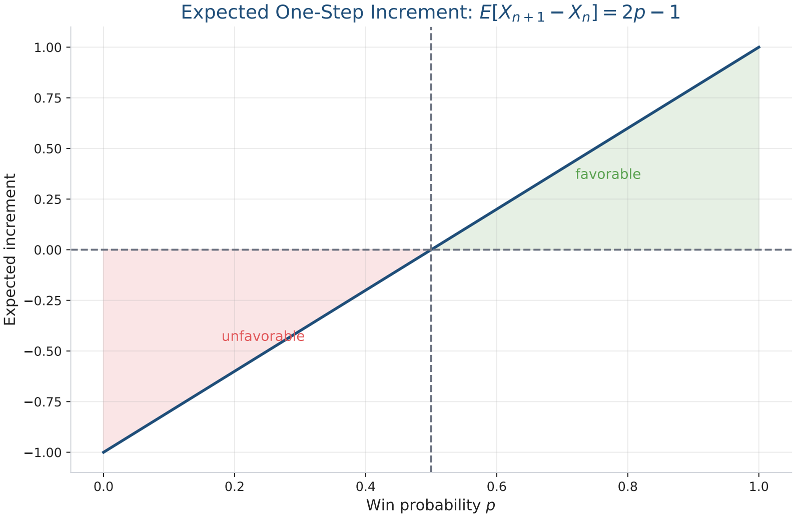
$$E[X_{n+1} - X_n] = (+1) \cdot p + (-1) \cdot q = p - q = 2p - 1.$$

- Therefore:
  - if $p < \frac{1}{2}$, the drift is negative,
  - if $p = \frac{1}{2}$, the drift is zero,
  - if $p > \frac{1}{2}$, the drift is positive.
- In roulette,

$$2p - 1 = 2 \cdot \frac{18}{37} - 1 = -\frac{1}{37} < 0.$$

- So repeated play tends to push the capital downward on average.

# Expected One-Step Increment



# Changing the Bet Size: Exact Values

So far we assumed that you always bet $\$1$.

If the bet size is $b$, then we measure capital in units of $b$:

$$i_b = \frac{10}{b}, \qquad K_b = \frac{80}{b}.$$

The same formula gives

$$R_b = \frac{\left(\frac{19}{18}\right)^{i_b} - \left(\frac{19}{18}\right)^{K_b}}{1 - \left(\frac{19}{18}\right)^{K_b}}.$$

| Bet size $b$ | Initial chips $i_b$ | Target $K_b$ | Ruin probability |
|---|---|---|---|
| $\$1$ | 10 | 80 | 0.9904 |
| $\$2$ | 5 | 40 | 0.9597 |
| $\$5$ | 2 | 16 | 0.9170 |
| $\$10$ | 1 | 8 | 0.8973 |

# Changing the Bet Size: Interpretation

- The $\$1$ betting strategy has

$$R_1 \approx 0.9904, \qquad 1 - R_1 \approx 0.0096.$$

- For larger bets:

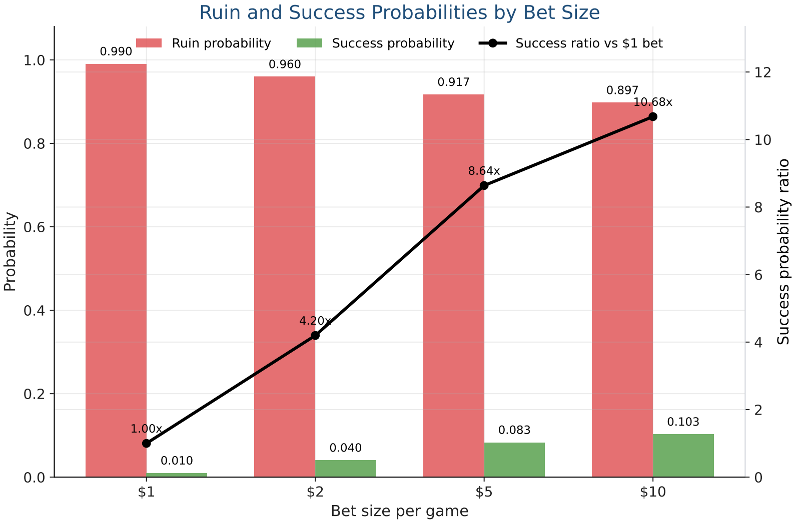
$$R_2 \approx 0.9597, \qquad R_5 \approx 0.9170, \qquad R_{10} \approx 0.8973.$$

- Compared with the $\$1$ strategy, the success probability becomes approximately

$$4.20\times, \quad 8.64\times, \quad 10.68\times$$

  larger for $\$2, \$5, \$10$ bets, respectively.
- The game is still unfavorable.
- However, larger bets reduce the number of games needed to reach $\$80$.

# Ruin Probability by Bet Size



# Can We Do Better?

So far, we considered **constant betting strategies**:

$$\$1, \quad \$2, \quad \$5, \quad \$10$$

per game.

- These strategies use the same bet size at every step.
- But the player may choose the bet depending on the current capital.
- This leads to a new question:

  Can a state-dependent strategy reduce the ruin probability?

- Since the game is unfavorable, a natural idea is:

  try to reach $\$80$ in as few games as possible.

# Bold Play: Idea

- **Bold play** is an extreme version of this idea.
- At each step, the player bets as much as possible, but does not overshoot the target.
- If the current capital is $i$ and the target is $80$, the bet is

$$b(i) = \min(i, 80 - i).$$

- Thus:
  - if $i \leq 40$, bet all current capital,
  - if $i > 40$, bet only $80 - i$, the amount needed to reach the target.

# Bold Play: First Example

Starting from $\$10$, bold play gives the path

$$10 \longrightarrow 20 \longrightarrow 40 \longrightarrow 80.$$

Thus success requires three consecutive wins:

$$P(\text{success}) = \left(\frac{18}{37}\right)^3 \approx 0.1151.$$

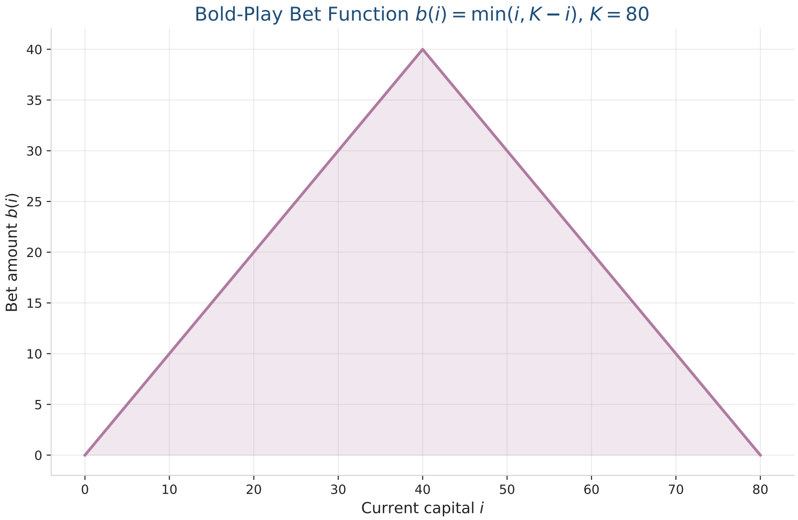
Therefore,

$$R_{\text{bold}}(10) = P(\text{ruin}) = 1 - \left(\frac{18}{37}\right)^3 \approx 0.8849.$$

**Comparison**

Bold play is still risky, but its success probability is much larger than the $\$1$-bet strategy.

# Bold-Play Betting Rule



# Bold Play: Recursive Formula

For a general capital $i$ and target $K$, bold play uses the bet

$$b(i) = \min(i, K - i).$$

Then the next state is

$$i + b(i) \quad \text{after a win,}$$

and

$$i - b(i) \quad \text{after a loss.}$$

Thus the ruin probability under bold play satisfies

$$R_{\text{bold}}(i) = pR_{\text{bold}}(i + b(i)) + (1 - p)R_{\text{bold}}(i - b(i)),$$

with boundary conditions

$$R_{\text{bold}}(0) = 1, \qquad R_{\text{bold}}(K) = 0.$$

# Bold Play: Explicit Piecewise Recursion

Since

$$b(i) = \min(i, K - i),$$

we obtain two cases.
If $i \leq K/2$, then $b(i) = i$. Hence

$$i + b(i) = 2i, \qquad i - b(i) = 0.$$

Therefore

$$R_{\text{bold}}(i) = pR_{\text{bold}}(2i) + (1 - p).$$

If $i > K/2$, then $b(i) = K - i$. Hence

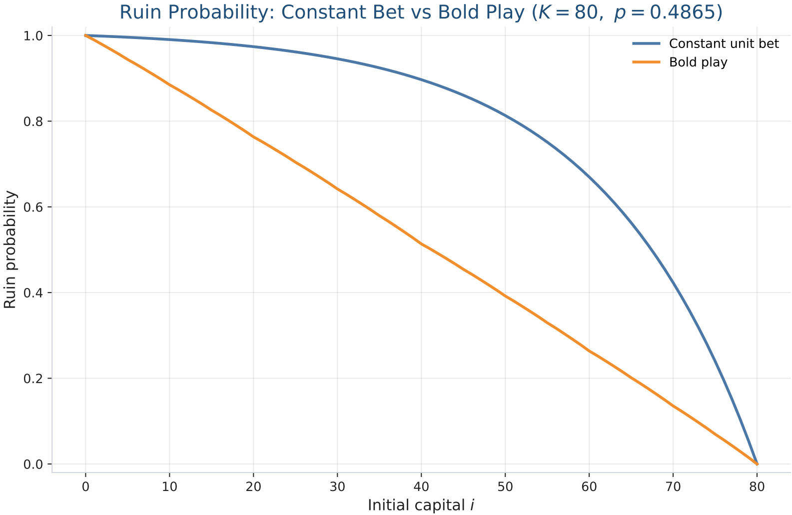
$$i + b(i) = K, \qquad i - b(i) = 2i - K.$$

Therefore

$$R_{\text{bold}}(i) = (1 - p)R_{\text{bold}}(2i - K).$$

This recursion explains the shape of the bold-play ruin probability curve.

# Constant Bet vs Bold Play



# Comparing Strategies

- The $\$1$ betting strategy gives

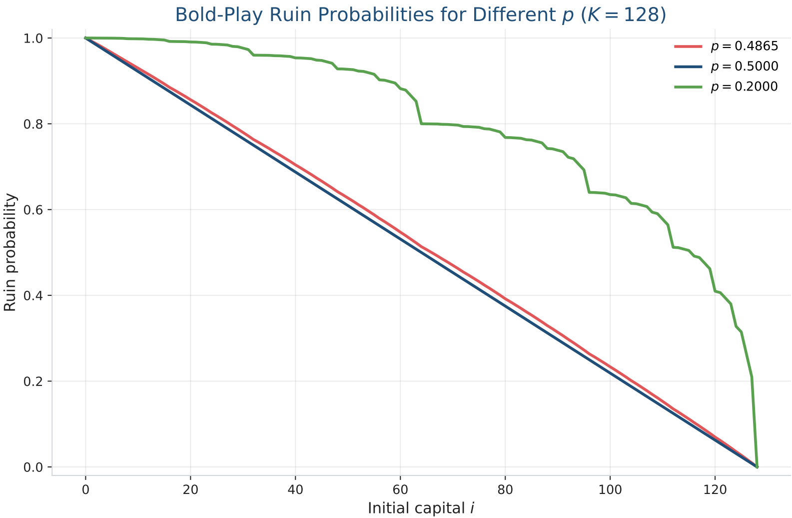
$$R_1(10) \approx 0.9904, \qquad 1 - R_1(10) \approx 0.0096.$$

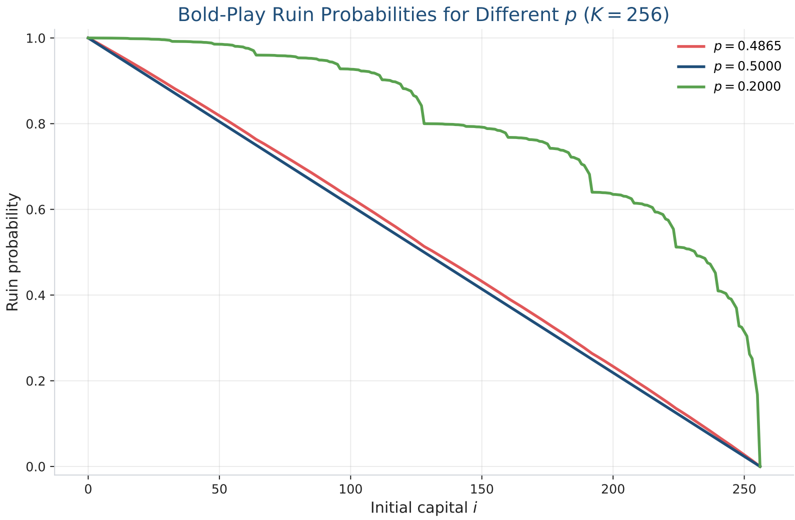
- Bold play gives

$$R_{\text{bold}}(10) \approx 0.8849, \qquad 1 - R_{\text{bold}}(10) \approx 0.1151.$$

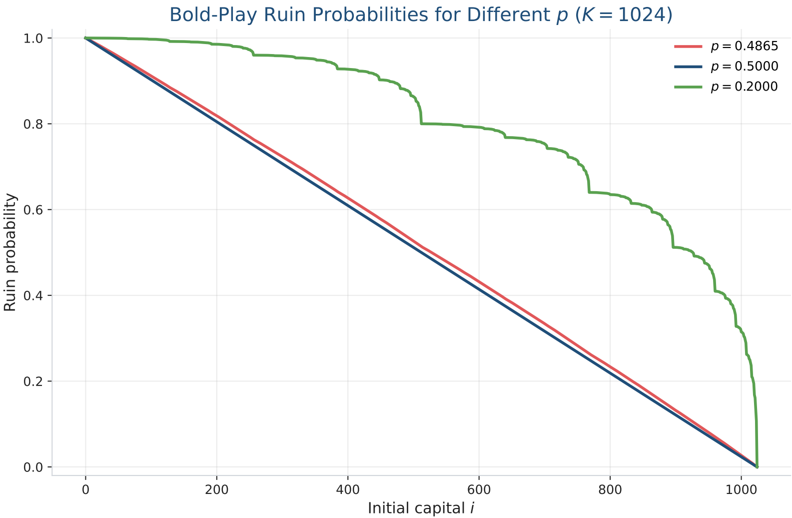
- So bold play is still very risky, but it gives a much larger success probability than betting $\$1$ each time.
- The reason is that in an unfavorable game, reducing the number of plays can be beneficial.

# Bold Play for $K = 128$

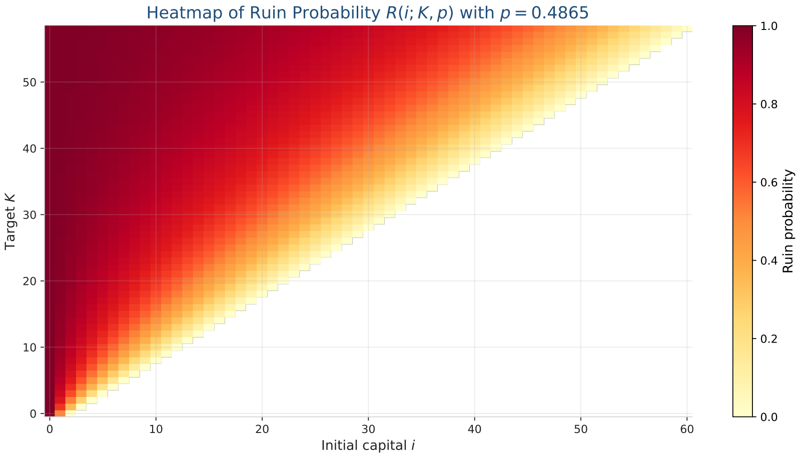



Bold-Play Ruin Probabilities for Different p (K = 256)
Ruin probability
Initial capital i
p = 0.4865
p = 0.5000
p = 0.2000
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250

# Bold Play for $K = 1024$



# Heatmap of Ruin Probabilities



# Interpretation in Insurance Mathematics

- The same mathematical idea appears in insurance risk modelling.
- Let

$$Z_n(i) = i + na - \sum_{k=1}^{n} S_k,$$

where

  - $i$: initial reserve,
  - $a$: premium income per period,
  - $S_k$: claim amount in period $k$.

- Ruin occurs when the reserve becomes non-positive:

$$P(\exists n \geq 1 : \; Z_n(i) \leq 0).$$

- So the casino story is a simple model that leads naturally to a much broader risk problem.

# Main Message of Part I

- We started from a concrete story:
  - you have $\$10$,
  - you want to make $\$80$,
  - you repeatedly bet on red.
- This story leads to:
  - a random walk,
  - a recursive equation,
  - an explicit probability formula,
  - a comparison of strategies,
  - and an interpretation in insurance mathematics.
- Main lesson:

A small disadvantage can accumulate dramatically over repeated play.

# Part II. The Monty Hall Problem

- We now turn to another classical probability example.
- This problem looks simple at first, but it teaches an important lesson:

  probability changes when information is revealed.

- We will explain the situation slowly, step by step.

# Step 1. The Basic Setting

- Imagine that you are in a TV game show.
- There are $3$ closed doors.
- Behind exactly one door there is a car.
- Behind the other two doors there are goats.
- Your goal is, of course, to get the car.

$$P(\text{car behind each door}) = \frac{1}{3}.$$

# Step 2. Your First Choice

- At the beginning, you do not know where the car is.
- So you choose one door.
- For convenience, let us say that you choose **door 1**.
- At this moment, the probability that your choice is correct is

$$\frac{1}{3}.$$

- The probability that your choice is wrong is

$$\frac{2}{3}.$$

# Step 3. The Host Knows the Answer

- Now the host acts.
- The host knows where the car is.
- The host follows these rules:
  - the host never opens the door you chose,
  - the host never opens the door with the car,
  - the host always opens a door with a goat.
- If the host has two possible goat doors to open, he chooses one of them.

This is the key point: the host's action is **not arbitrary**.

# Step 4. What Happens Next?

- Suppose you chose door $1$.
- Then the host opens one of the other two doors.
- The opened door always shows a goat.
- So after the host acts, only two unopened doors remain:
  - your original door,
  - one other closed door.

Now you must decide whether to

- **stay** with your original choice, or
- **switch** to the other unopened door.

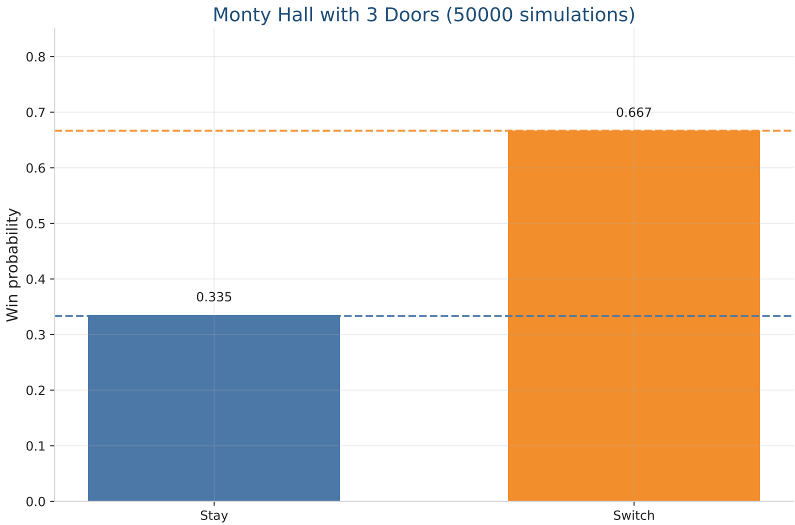
# Step 5. A Concrete Example

- Let us keep the same setting:
  - you chose door $1$,
  - then the host opens door $3$,
  - behind door $3$ there is a goat.
- So now only doors $1$ and $2$ remain closed.

Question:

Should you stay with door $1$, or switch to door $2$?

# Monty Hall: Stay vs Switch



## You should switch.

- Probability of winning by staying:
$$\frac{1}{3}$$
- Probability of winning by switching:
$$\frac{2}{3}$$

# Step 7. Why Is Staying Only $\frac{1}{3}$?

- At the beginning, your first choice was correct with probability

$$\frac{1}{3}.$$

- If you decide to **stay**, you are keeping exactly that first choice.
- So the probability of winning by staying is still

$$\frac{1}{3}.$$

The host's action does not make your original choice better.

# Step 8. Why Is Switching $\frac{2}{3}$?

- Your first choice was wrong with probability

$$\frac{2}{3}.$$

- If your first choice is wrong, then the car must be behind one of the other two doors.
- But the host opens one goat door.
- Therefore the only remaining closed door must contain the car.

So switching wins exactly when your first choice was wrong.

$$P(\text{win by switching}) = \frac{2}{3}.$$

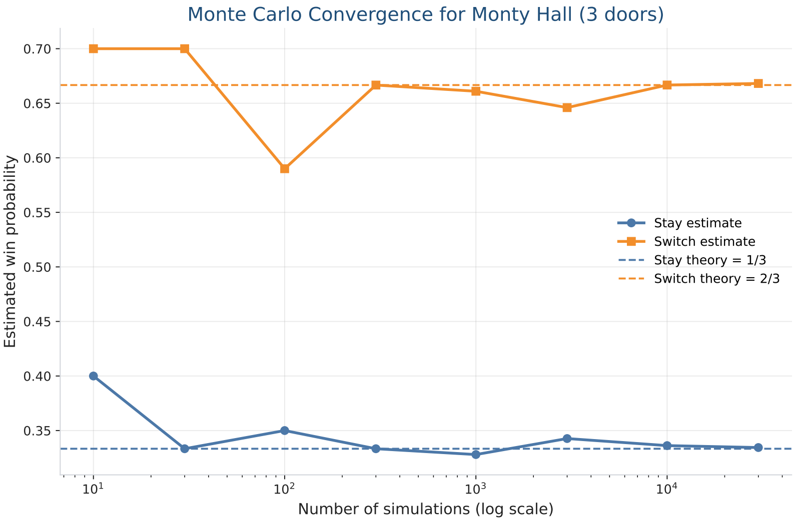
# Step 9. The Main Idea in One Sentence

**Key idea**

The probability $\frac{2}{3}$ does not disappear. It moves to the other unopened door after the host reveals a goat.

- Your chosen door starts with probability $\frac{1}{3}$.
- The other two doors together start with probability $\frac{2}{3}$.
- The host removes one goat door from that $\frac{2}{3}$ group.
- So the remaining unopened door keeps probability $\frac{2}{3}$.

# Monte Carlo Convergence for the 3-Door Problem



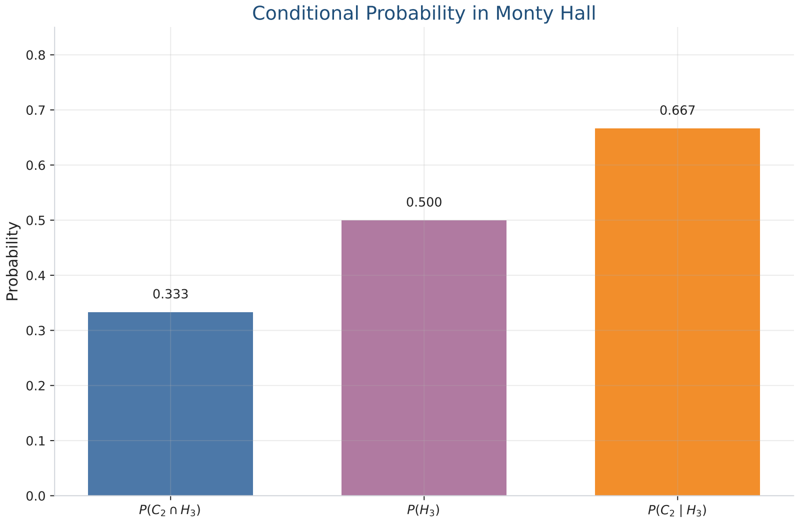
# Step 10. A Conditional Probability Question

- Let us now write the same idea more mathematically.
- Suppose:
  - you first choose door $1$,
  - the host opens door $3$.
- Then we ask:

$$P(\text{car behind door } 2 \mid \text{host opens door } 3).$$

- This is a conditional probability.

# A Conditional Probability Calculation



# Step 11. Why It Is Not $\frac{1}{2}$

- Some people say:

  "*Now two doors remain, so each should have probability* $\frac{1}{2}$*.*"

- This is the common mistake.
- The two remaining doors are **not symmetric**, because the host used information.
- The host:
  - knows where the car is,
  - never opens the car door,
  - never opens your chosen door.
- Therefore the host's action changes the conditional probabilities.

# Step 12. A Simple Summary of the 3-Door Case

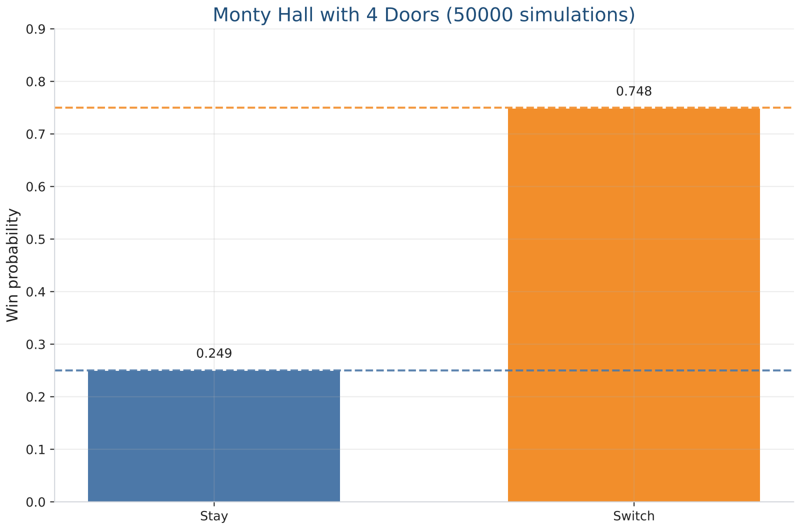
- If your first choice is correct, then switching loses.
- If your first choice is wrong, then switching wins.
- Since

$$P(\text{first choice correct}) = \frac{1}{3}, \qquad P(\text{first choice wrong}) = \frac{2}{3},$$

we get

$$P(\text{win by stay}) = \frac{1}{3}, \qquad P(\text{win by switch}) = \frac{2}{3}.$$

# The 4-Door Version



# Step 13. The 4-Door Version

- Suppose now there are $4$ doors.
- One door has a car, and the other three have goats.
- You choose one door first.
- Then the host opens **two** goat doors.
- Only your original door and one other closed door remain.

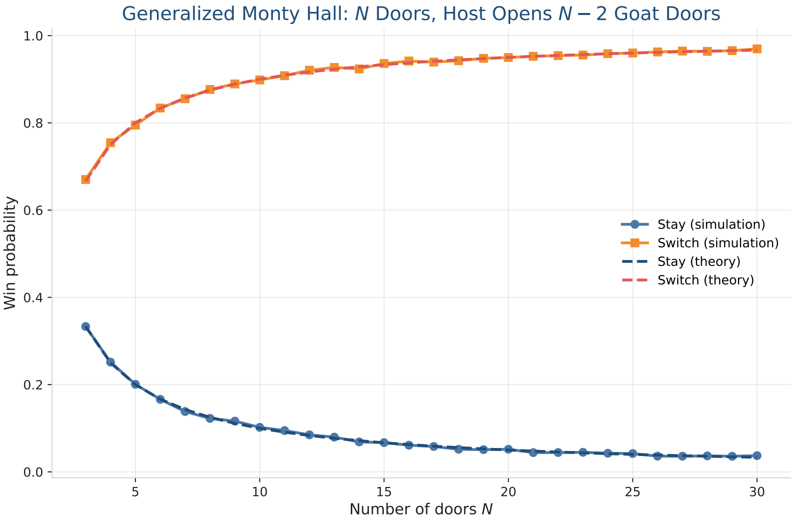
Now:

$$P(\text{first choice correct}) = \frac{1}{4}, \qquad P(\text{first choice wrong}) = \frac{3}{4}.$$

So switching wins with probability

$$\frac{3}{4}.$$

# Generalized Monty Hall with $N$ Doors



# Step 14. The $N$-Door Version

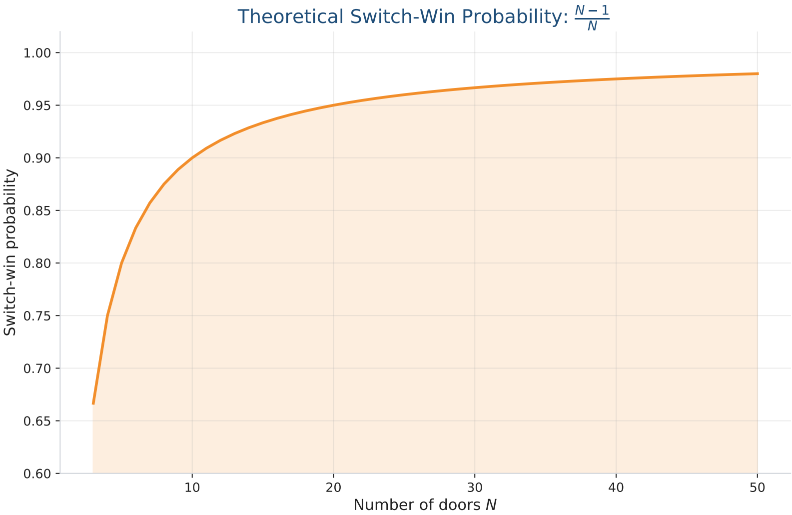
- Suppose there are $N$ doors.
- One door has a car, and the other $N-1$ doors have goats.
- You choose one door first.
- Then the host opens $N-2$ goat doors.
- Only two closed doors remain.

Then:

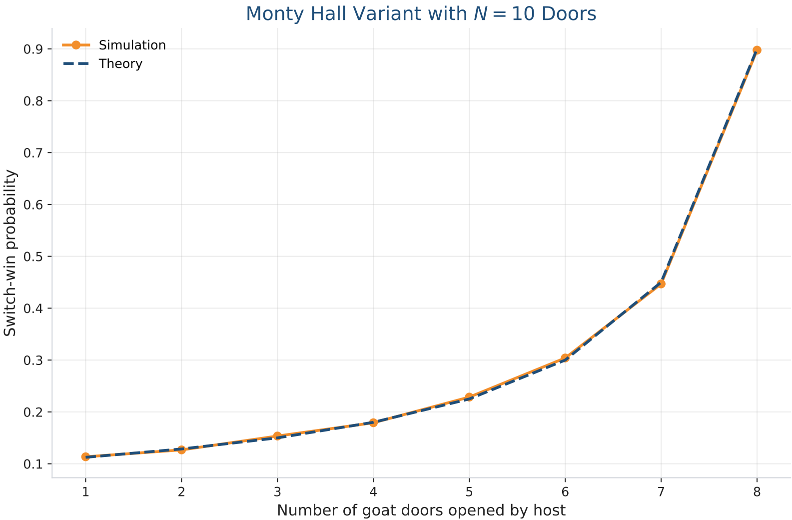
$$P(\text{win by staying}) = \frac{1}{N}, \qquad P(\text{win by switching}) = \frac{N-1}{N}.$$

So the larger $N$ becomes, the more favorable switching becomes.

# Theoretical Switch-Win Probability



# Variant: The Host Opens Only Some Goat Doors



# Step 15. A Variant

- There is also a variant.
- Suppose there are $N$ doors, but now the host opens only $p$ goat doors, not all $N - 2$ of them.
- Then more than two closed doors remain.
- If you switch uniformly to one of the remaining closed doors, the winning probability changes.

So the answer depends on

- the number of doors,
- and how much information the host reveals.

# Main Message of Part II

- The Monty Hall problem is not only a puzzle.
- It teaches an important lesson:

probability must be updated when new information is revealed.

- The host's action is informative.
- That is why conditional probability is essential here.

# Why Do These Examples Matter?

- **Ruin probability** is not only a gambling problem.
  - It is a basic model for insurance risk, reserve management,
  - and reliability under repeated random losses.
- **Monty Hall** is not only a puzzle.
  - It shows how probabilities change when new information is revealed,
  - which is a key idea in statistics, Bayesian inference and data-driven decision making.
- More broadly, probabilistic modelling appears in:
  - finance and economics,
  - medicine and biology,
  - engineering and control,
  - machine learning and artificial intelligence.

# Closing

**Thank you.**