

# Monty Hall Probability Plot

Kyle Larson

2025-08-11

## Introduction

Calculate each of the probabilities, for sample sizes 4 to 40 of binary pass/fail samples, that the majority of the samples in each size is a failure knowing that there is a %33 chance of failure and that - for each- 3 have already been determined failures. Assume pass fail outcome is independent of one sample from another. Does this probability increase or decrease ?

This will generate a plot showing how the probability of majority failures behaves as sample size increases from 4 to 40.

You'll likely see that the probability decreases as sample size increases, since the 3 known failures have less influence in larger samples.

```
# Load required library
library(ggplot2)

# Parameters
p_fail <- 0.33
n_values <- 4:40
prob_majority_failures <- numeric(length(n_values))

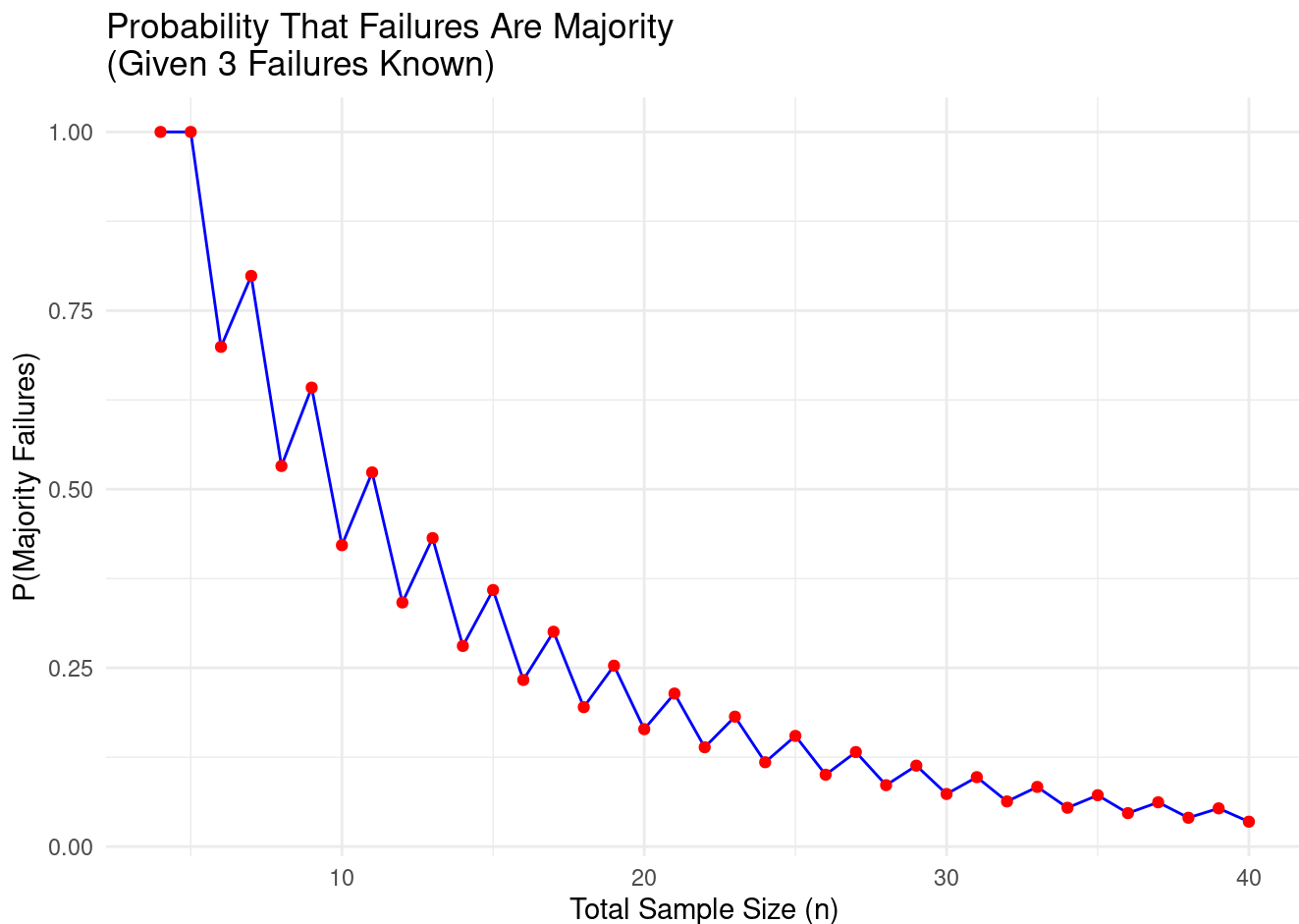
# Loop over sample sizes
for (i in seq_along(n_values)) {
  n <- n_values[i]
  remaining_samples <- n - 3
  threshold <- (n / 2) - 3
  k <- floor(threshold) + 1

  # Compute  $P(X \geq k)$  for  $X \sim \text{Binomial}(\text{remaining\_samples}, 0.33)$ 
  x_vals <- k:remaining_samples
  prob <- sum(dbinom(x_vals, size = remaining_samples, prob = p_fail))

  prob_majority_failures[i] <- prob
}

# Create a data frame for plotting
df <- data.frame(
  SampleSize = n_values,
  Probability = prob_majority_failures
)

# Plot the results
ggplot(df, aes(x = SampleSize, y = Probability)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +
  theme_minimal() +
  labs(
    title = "Probability That Failures Are Majority\n(Given 3 Failures Known)",
    x = "Total Sample Size (n)",
    y = "P(Majority Failures)"
  )
}
```



## Interpretation

The point at which the probability that the majority is failure drops below 5% at 40

Rather, given 3 failures, the probability that at least half are also rejects when sample size is 11 is %50

```
print(prob_majority_failures)
```

```
## [1] 1.00000000 1.00000000 0.69923700 0.79848879 0.53249399 0.64221685
## [7] 0.42167391 0.52356475 0.34151979 0.43163204 0.28069401 0.35896616
## [13] 0.23310455 0.30059784 0.19508333 0.25311631 0.16424294 0.21410090
## [19] 0.13894003 0.18178172 0.11799520 0.15483192 0.10053459 0.13223588
## [25] 0.08589436 0.11320275 0.07356007 0.09710783 0.06312661 0.08345162
## [31] 0.05427071 0.07183075 0.04673159 0.06191658 0.04029695 0.05343941
## [37] 0.03479258
```

## Monty Hall

The Monty Hall problem is a probability puzzle based on a game show scenario with three doors: behind one is a prize, and behind the others are goats. The player picks a door, then the host—who knows what's behind each door—opens one of the two remaining doors to reveal a goat. The player is then given the option to switch doors, and counterintuitively, switching increases the chance of winning from  $1/3$  to  $2/3$ .

```
# Load necessary library
library(ggplot2)

# Range of doors from 3 to 33
door_counts <- 3:33

# Probability of winning if switching:
# With n doors, the player initially has 1/n chance of picking the prize.
# If they switch, the chance of winning becomes (n - 1)/n * 1/(n - 2)
# (Host opens n - 2 doors revealing goats, leaving one unchosen unopened door)

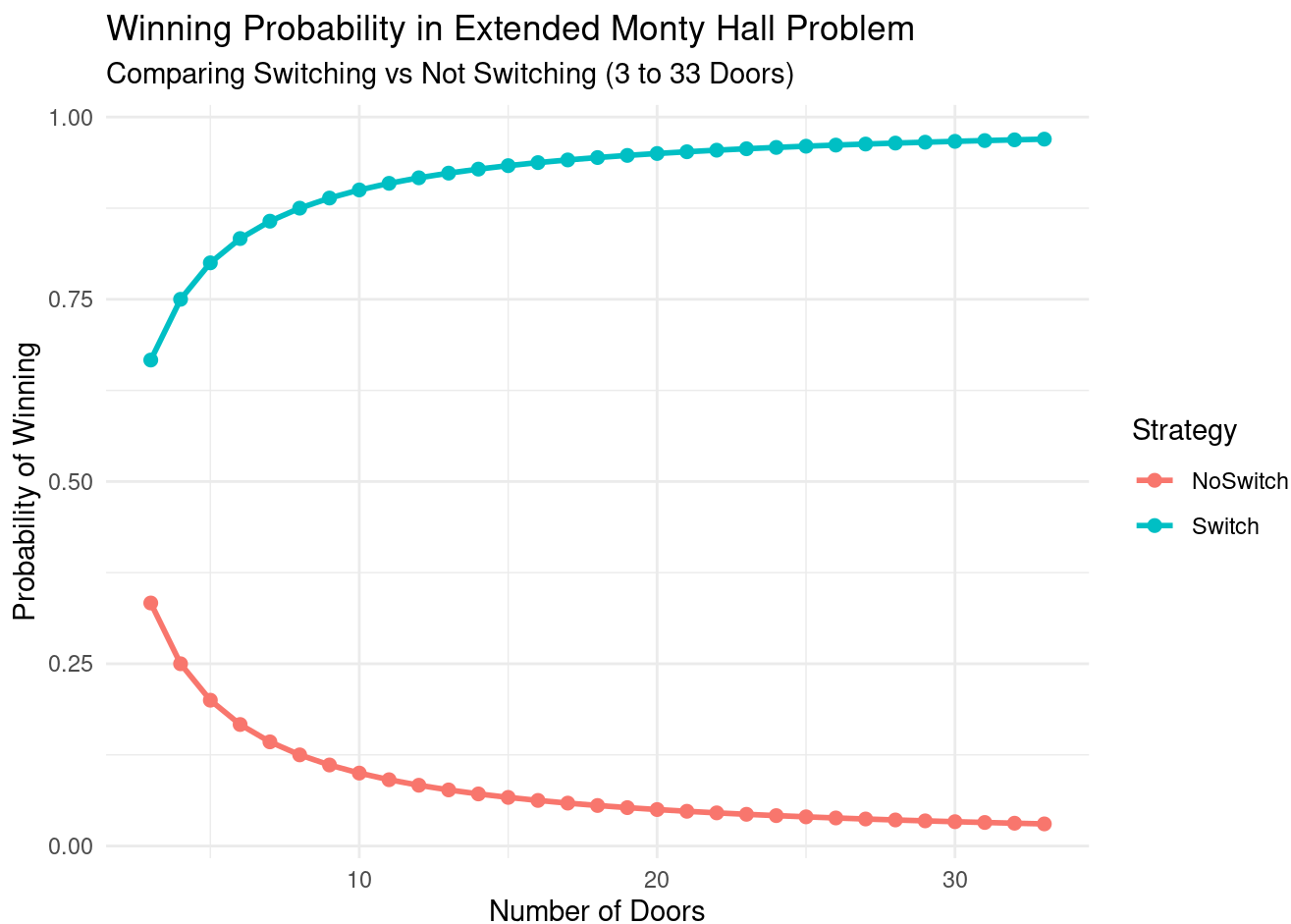
prob_switch <- (door_counts - 1) / door_counts
prob_noswitch <- rep(1 / door_counts)

# Create data frame for plotting
df <- data.frame(
  Doors = door_counts,
  Switch = prob_switch,
  NoSwitch = prob_noswitch
)

# Reshape data for ggplot
library(tidyr)
df_long <- pivot_longer(df, cols = c("Switch", "NoSwitch"),
                        names_to = "Strategy", values_to = "Probability")

# Plotting the probabilities
ggplot(df_long, aes(x = Doors, y = Probability, color = Strategy)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(
    title = "Winning Probability in Extended Monty Hall Problem",
    subtitle = "Comparing Switching vs Not Switching (3 to 33 Doors)",
    x = "Number of Doors",
    y = "Probability of Winning",
    color = "Strategy"
  )
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



To state this again - the probability of winning after switching doors is  $(1 - (1/n))$  the probability that your initial assessment was incorrect all multiplied by the probability of success knowing that an additional remaining  $n-2$  doors (shown by the knowing host "Monty") having taken one away from the original  $n-1$ . That is you have 1 in  $n-2$  times the probability that your initial guess was incorrect.