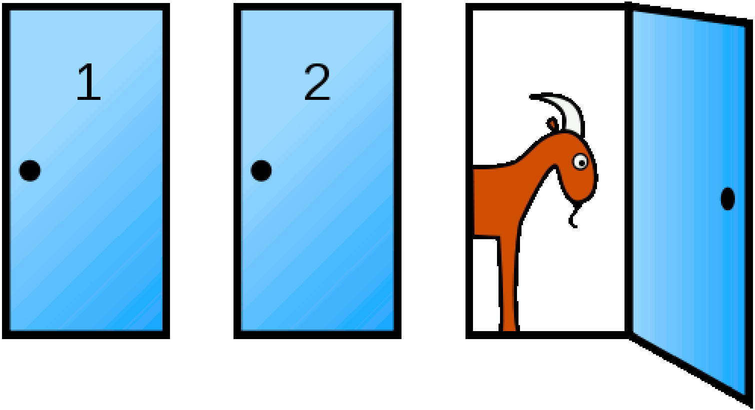
Truth is stranger than fiction: A look at some improbabilities

Rick Durrett


Monty Hall!

You picked door 1, should you switch?



Marilyn vos Savant is an American magazine columnist, author, lecturer and playwright who rose to fame through her listing in the Guinness Book of World Records under "Highest IQ". Since 1986 she has written Ask Marilyn, a Sunday column in Parade magazine in which she solves puzzles and answers questions from readers on a variety of subjects.

Her Sept. 9, 1990 column was devoted to the Monty Hall problem. Vos Savant answered arguing that the selection should be switched to door #2 because it has a 2/3 chance of success, while door #1 has just 1/3.

Reaction to Marilyn vos Savant's

You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame!

Scott Smith, Ph.D.
University of Florida

See marilynvossavant.com for the original column and many of the letters.

Solution to Monty Hall

Suppose $\#1$ is chosen.

	$\#1$	$\#2$	$\#3$	host's action
case 1	donkey	donkey	car	opens $\#2$
case 2	donkey	car	donkey	opens $\#3$
case 3	car	donkey	donkey	opens $\#2$ or $\#3$

$P(\text{case } 2, \text{open door } \#3) = 1/3$ and

$$P(\text{case } 3, \text{open door } \#3) = P(\text{case } 3)P(\text{open door } \#3|\text{case } 3) = \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}$$

$P(\text{open door } \#3) = 1/3 + 1/6 = 1/2$ so

$$P(\text{case } 3|\text{open door } \#3) = \frac{P(\text{case } 3, \text{open door } \#3)}{P(\text{open door } \#3)} = \frac{1/6}{1/2} = \frac{1}{3}$$

Easier Solution

Your probability of winning was $1/3$ when you picked and it didn't change when Monty opened door 3.

Cognitive Dissonance in Monkeys

Yale psychologists measured monkeys preferences by observing how quickly each monkey sought out different colors of M&Ms. In the first step, the researchers gave the monkey a choice between say red and blue. If the monkey chose red, then it was given a choice between blue and green. **Nearly two-thirds of the time** it rejected blue in favor of green, which seemed to jibe with the theory of choice rationalization:

"once we reject something, we tell ourselves we never liked it anyway."

Who's the monkey?

There are six possible orderings:

RGB	BRG	GRB
RBG	BGR	GBR

In three of these (in red) $R > G$ and in $2/3$'s of these $B > G$.

Observation of economist M. Keith Chen.

Story from New York Times, April 2008.

Birthday problem

Someone wants to bet you \$20 that in a group of 40 people (e.g., the Boston Red Sox roster of active players) no two have the same birthday (day only not year). Should you take the bet?

Birthday problem

Someone wants to bet you \$20 that in a group of 40 people (e.g., the Boston Red Sox roster of active players) no two have the same birthday (day only not year). Should you take the bet?

A match is likely because there are $(40 \cdot 39)/2 = 780$ pairs of players that share a birthday with probability $1/365$. Probability of no match is about 0.11.

Birthday problem

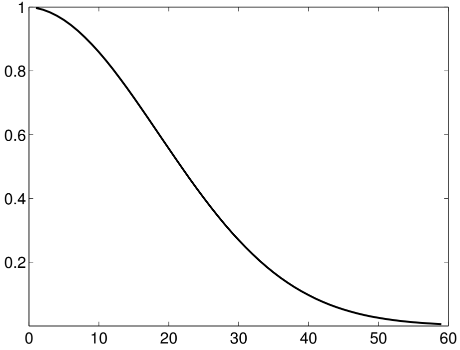
Someone wants to bet you \$20 that in a group of 40 people (e.g., the Boston Red Sox roster of active players) no two have the same birthday (day only not year). Should you take the bet?

A match is likely because there are $(40 \cdot 39)/2 = 780$ pairs of players that share a birthday with probability $1/365$. Probability of no match is about 0.11.

On the Atlanta Braves 2014 roster Craig Kimbrel and Andy Northcraft were both born on May 28, and Dan Uggla and Jordan Schaefer were born on September 4.

Probability all birthdays different for n people

$$\frac{365 \cdot 364 \cdots 366 - n}{(365)^n}$$



P(triple birthday in the Senate) = ?

Challenge: calculate the probability that in a group of 100 people there is at least one triple birthday.

Approximate solution: The number of senators born on a given day is Binomial(100,1/365) and hence approximately Poisson with mean 100/365.

Probability of triple $= e^{-100/365}(100/365)^3/3! = 0.002606$, expected number among 365 days $= 0.9512$

Probability of double $= e^{-100/365}(100/365)^2/2! = 0.028536$, expected number among 365 days $= 10.4$

Answers for this Senate (2018)

One triple birthday May 3: Jim Risch (Idaho), David Vitter (Louisiana), Ron Wyder (Oregon)

Ten double birthdays: Jan 7, March 31, June 22, August 24, Sept 29, October 20 and 24, November 17, December 7 and 10

Diane Feinstein and Elizabeth Warren were both born on June 22

A Birthday Triple

Chad and Barbara Soper will have an easy time remembering their kids birthdays since they are

$$8/8/8 \qquad 9/9/9 \qquad 10/10/10$$

A Birthday Triple

Chad and Barbara Soper will have an easy time remembering their kids birthdays since they are

$$8/8/8 \qquad 9/9/9 \qquad 10/10/10$$

Philip Stark, Statistics U.C. Berkeley: "The chance you get depends on your assmptions. One set of assumptions gives you a chance of about 1 in 50 million (which is $1/365^3$).

More realistic assumptions allowing for at least 11 months between births increases the probability to about 1/2500."

A Birthday Triple

Chad and Barbara Soper will have an easy time remembering their kids birthdays since they are

$$8/8/8 \qquad 9/9/9 \qquad 10/10/10$$

Philip Stark, Statistics U.C. Berkeley: "The chance you get depends on your assmptions. One set of assumptions gives you a chance of about 1 in 50 million (which is $1/365^3$).

More realistic assumptions allowing for at least 11 months between births increases the probability to about 1/2500."

Complication. Labor was induced for the third child

A Birthday Triple

Chad and Barbara Soper will have an easy time remembering their kids birthdays since they are

8/8/8   9/9/9   10/10/10

Philip Stark, Statistics U.C. Berkeley: "The chance you get depends on your assmptions. One set of assumptions gives you a chance of about 1 in 50 million (which is $1/365^3$).

More realistic assumptions allowing for at least 11 months between births increases the probability to about 1/2500."

Complication. Labor was induced for the third child on 10/9/10.

Pick 4 coincidence

To quote a United Press story on September 10, 1981:

> *"Lottery officials say that there is 1 chance in 100 million (10^8) that the same four digit lottery number would be drawn in Massachusetts and New York on the same night. That's just what happened Tuesday. The number 8902 came up paying \$5842 in Massachusetts and \$4500 New York."*

Pick 4 coincidence

To quote a United Press story on September 10, 1981:

> *"Lottery officials say that there is 1 chance in 100 million (10^8) that the same four digit lottery number would be drawn in Massachusetts and New York on the same night. That's just what happened Tuesday. The number 8902 came up paying $5842 in Massachusetts and $4500 New York."*

Some number will be picked in MA.
NY will match with probability 10^{-4}

Sally Clark

In 1999, a British jury convicted Sally Clark of murdering her two children who had died suddenly at the ages of 11 and 8 weeks respectively of sudden infant death syndrome or "cot deaths". There was no physical or other evidence of a murder, nor was there a motive. Most likely the jury was convinced by a pediatrician who said that a baby had a probability of roughly 1/8500 of dying a cot death, so having two children die this way had probability roughly 1/73,000,000.

Sally Clark

In 1999, a British jury convicted Sally Clark of murdering her two children who had died suddenly at the ages of 11 and 8 weeks respectively of sudden infant death syndrome or "cot deaths". There was no physical or other evidence of a murder, nor was there a motive. Most likely the jury was convinced by a pediatrician who said that a baby had a probability of roughly 1/8500 of dying a cot death, so having two children die this way had probability roughly 1/73,000,000.

$\rightarrow$ 1/8500 many people experience one cot death

Sally Clark

In 1999, a British jury convicted Sally Clark of murdering her two children who had died suddenly at the ages of 11 and 8 weeks respectively of sudden infant death syndrome or "cot deaths". There was no physical or other evidence of a murder, nor was there a motive. Most likely the jury was convinced by a pediatrician who said that a baby had a probability of roughly 1/8500 of dying a cot death, so having two children die this way had probability roughly 1/73,000,000.

$\rightarrow$ 1/8500 many people experience one cot death

$\rightarrow$ 1/100 one one child has this, the risk for others increases

Sally Clark

In 1999, a British jury convicted Sally Clark of murdering her two children who had died suddenly at the ages of 11 and 8 weeks respectively of sudden infant death syndrome or "cot deaths". There was no physical or other evidence of a murder, nor was there a motive. Most likely the jury was convinced by a pediatrician who said that a baby had a probability of roughly 1/8500 of dying a cot death, so having two children die this way had probability roughly 1/73,000,000.

$\rightarrow$ 1/8500 many people experience one cot death

$\rightarrow$ 1/100 one one child has this, the risk for others increases

Sally Clark spent 3 years in jail before the conviction was overturned

Lottery Double Winner

A New Jersey woman, Evelyn Adams, won the lottery twice within a span of four months raking in a total of 5.4 million dollars. She won the jackpot for the first time on October 23, 1985 in the Lotto 6/39 in which you pick 6 numbers out of 39. Then she won the jackpot in the new Lotto 6/42 on February 13, 1986. Lottery officials calculated the probability of this as roughly one in 17.1 trillion.

$$\frac{1}{C_{39,6}} \cdot \frac{1}{C_{42,6}} = \frac{1}{17.1 \times 10^{12}}$$

$C_{n,m} = n!/m!(n-m)!$ is the number of ways of picking m things out of n.

What's wrong with this calculation?

Somebody won the October 23, 1985 lottery.

What's wrong with this calculation?

Somebody won the October 23, 1985 lottery.

We would have been equally impressed if this happened twice within a one year period. (100 twice weekly drawings)

What's wrong with this calculation?

Somebody won the October 23, 1985 lottery.

We would have been equally impressed if this happened twice within a one year period. (100 twice weekly drawings)

Many people who play the lottery buy more than one ticket. Suppose 1,000,000 people buy 5 tickets each.

What's wrong with this calculation?

Somebody won the October 23, 1985 lottery.

We would have been equally impressed if this happened twice within a one year period. (100 twice weekly drawings)

Many people who play the lottery buy more than one ticket. Suppose 1,000,000 people buy 5 tickets each.

Probability is now about $1/200$. Now take into account the number of states with lotteries.

Maureen Wilcox.

In June 1980 she bought tickets for both the Massachusetts Lottery and the Rhode Island lottery. She picked the winning numbers for both lotteries.

Maureen Wilcox.

In June 1980 she bought tickets for both the Massachusetts Lottery and the Rhode Island lottery. She picked the winning numbers for both lotteries.

Unfortunately for her, her Massachusetts numbers won in Rhode Island and vice versa.

Scratch-off Triple Winner.

81-year old Keith Selix won three lottery prizes totaling \$81,000 from scratch off games in the twelve months preceding May 3, 2006. He won \$30,000 twice in "Wild Crossword" games and \$21,000 playing "Double Blackjack." The odds of winning in these games are 89,775 to 1 and 119,700 to 1 respectively.

Scratch-off Triple Winner.

81-year old Keith Selix won three lottery prizes totaling \$81,000 from scratch off games in the twelve months preceding May 3, 2006. He won \$30,000 twice in "Wild Crossword" games and \$21,000 playing "Double Blackjack." The odds of winning in these games are 89,775 to 1 and 119,700 to 1 respectively.

One of the reasons Selix won so many times in 2006 is that he spent about \$200 a week or more than \$10,000 a year on scratch-off games. Expected number of wins $= 10^4/10^5 = 0.1$, so the probability of exactly three wins would be

$$e^{-0.1}\frac{(0.1)^3}{3!} \quad \text{or} \quad < \frac{1}{60,000}$$

Luckiest Gas Station in America (Bishop, TX)

Scratch-off Quadruple Winner

Joan Ginther, 63, has won \$21 million dollars since 1993 playing the lottery. Associated Press wrote that her chances of doing this were 1 in 18×10^{24} (18 septillion)

Scratch-off Quadruple Winner

Joan Ginther, 63, has won \$21 million dollars since 1993 playing the lottery. Associated Press wrote that her chances of doing this were 1 in 18×10^{24} (18 septillion)

Joan got her Ph.D. in Statistics from Stanford. An article in Harper's suggested that she may have figured out the random number generator that makes the tickets.

Ginther lives in Las Vegas but spends two months a year in Bishop Texas playing scratch off tickets.

The lucky store was shut down by the IRS once they discovered the store owner was holding boxes of high stakes scratch off tickets for her.

Too many winners

Powerball officials were amazed when 110 players won the second prize by getting five of the six numbers right in the March 30, 2005 drawing. Based on the number of tickets sold there should have only been an average of 4 winners. Since winning is a rare event the number of winners should be Poisson and the probability of 110 would be

$$e^{-4}\frac{4^{110}}{110!} \approx 10^{-110}$$

where we have used Stirling's formula $n! \sim (n/e)^n\sqrt{2\pi n}$.

Too many winners

Powerball officials were amazed when 110 players won the second prize by getting five of the six numbers right in the March 30, 2005 drawing. Based on the number of tickets sold there should have only been an average of 4 winners. Since winning is a rare event the number of winners should be Poisson and the probability of 110 would be

$$e^{-4}\frac{4^{110}}{110!} \approx 10^{-110}$$

where we have used Stirling's formula $n! \sim (n/e)^n\sqrt{2\pi n}$.

All 110 winners played the same six numbers: 22, 28, 32, 33, 39, 40*

Too many winners

Powerball officials were amazed when 110 players won the second prize by getting five of the six numbers right in the March 30, 2005 drawing. Based on the number of tickets sold there should have only been an average of 4 winners. Since winning is a rare event the number of winners should be Poisson and the probability of 110 would be

$$e^{-4}\frac{4^{110}}{110!} \approx 10^{-110}$$

where we have used Stirling's formula $n! \sim (n/e)^n\sqrt{2\pi n}$.

All 110 winners played the same six numbers: 22, 28, 32, 33, 39, 40*

The numbers came from fortune cookies from the same factory in Long Island City, Queens

Some people are too lucky

In Floridas Play 4 game you pick a four digit number like 3782 and if all four digits match you win $5000. Some people however are very good at winning this gamble. A recent paper on the arXiv:1503.02902v1 by Rich Arratia, Skip Garibaldi, Lawrence Mower, and Philip B. Stark says that an individual that they call LJ has won 57 times.

Now that by itself is not proof of guilt. If he bought 570,000 tickets he would end up with about this many wins. However that seems a little unlikely. If he only bought 250,000 tickets the probability of 57 wins is $1.22x10^{-8}$.

How did he get so lucky ?

There are three common schemes.

(i) A clerk can scratch an unsold ticket with a pin revealing enough of the bar code to be able to scan it to see if it is a winner.

(ii) Sometimes a customer will ask the clerk if the ticket was a winner. If so the clerk may lie about the ticket being a winner and keep the money himself.

(iii) Sometimes the winner may be an illegal immigrant or owe child support or back taxes, and will sell the ticket to an aggregator who pays half price for it and later claims the prize. This is a good scheme for people who want to launder money.

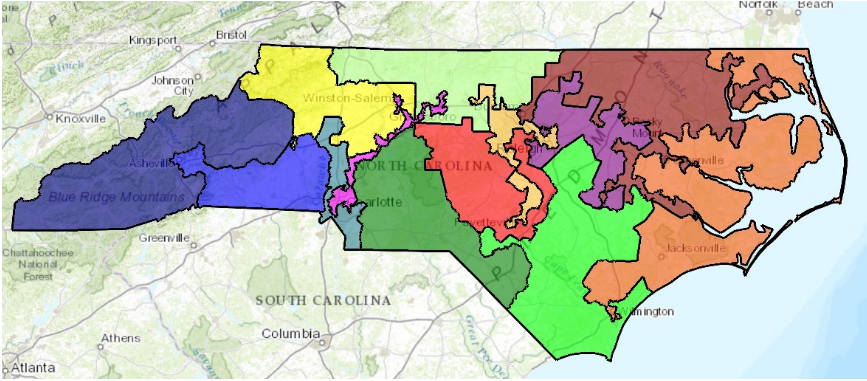
Gerrymandering

Suppose that a state has $200n$ Democrats and $200n$ Republicans. If there are four districts and we divide voters

District	Rep.	Dem.
1	$60n$	$40n$
2	$60n$	$40n$
3	$60n$	$40n$
4	$20n$	$80n$

then Republicans will win in 3 districts out of 4. Something like this was done in North Carolina resulting in 10 Republicans and 3 Democrats wins in 13 districts in the 2016 election, even though the split between the parties is roughly 50-50.

Weird shapes of North Carolina districts



Randomly drawn districts

To look for evidence of gerrymandering we can see what the outcomes would be if districts were drawn "at random." We can't just put the people of North Carolina into 13 districts at random. The districts should be

connected and compact

minimize the splitting of counties.

Jonathan Mattingly at Duke, working with undergrads, grad student, and postdocs has developed methods (see arXiv:1801.03783) for randomly sampling from possible redistricting plans. In an analysis they generate about 24,000 of these and retally the votes.

Results for North Carolina

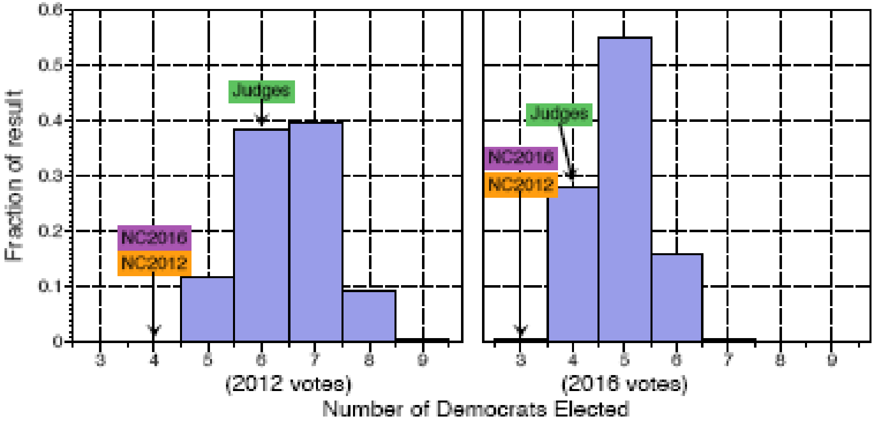


Figure: Outcomes as extreme as those observed in the election happen for fewer than 1% of redistricting plans. Obama won state in 2012, Trump in 2016.

The Gerrymandering work has been involved in a number of court cases in North Carolina, Pennsylvania, Wisconsin,
https://sites.duke.edu/quantifyinggerrymandering/

Slides can be found on my web page
https://services.math.duke.edu/~rtd/ Click on **Talks**