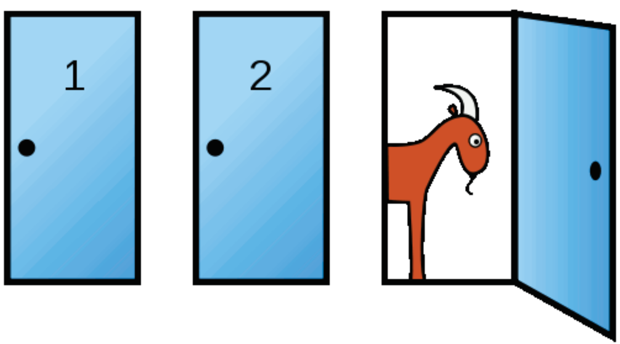
Truth is stranger than fiction: A look at some improbabilities

Rick Durrett

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	*GRB*	*BRG*
RBG	*GBR*	*BGR*

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

Observation of economist M. Keith Chen.

Story from New York Times, April 2008.

Another example of Mr. Bayes' formula

A 50 year old woman just received a call from her doctor saying there was suspicious spot on her mammogram. What is the probability she has breast cancer?

Another example of Mr. Bayes' formula

A 50 year old woman just received a call from her doctor saying there was suspicious spot on her mammogram. What is the probability she has breast cancer?

In a survey of 95 physicians the average answer was 75%.

Another example of Mr. Bayes' formula

A 50 year old woman just received a call from her doctor saying there was suspicious spot on her mammogram. What is the probability she has breast cancer?

In a survey of 95 physicians the average answer was 75%.

Suppose that 1% of women aged 50 have breast cancer, but there is a 5% chance of a false positive.

Another example of Mr. Bayes' formula

A 50 year old woman just received a call from her doctor saying there was suspicious spot on her mammogram. What is the probability she has breast cancer?

In a survey of 95 physicians the average answer was 75%.

Suppose that 1% of women aged 50 have breast cancer, but there is a 5% chance of a false positive.

In a group of 1000 women 10 will have breast cancer but $990 \cdot 0.05 \approx 50$ will have false positives, so the posterior probability is $10/60 = 16\%$. How if the disease prevalence is 0.1% (spina bifida) the answer is $1/51 = 2\%$.

Birthday problem

Someone wants to be you $20 that in a group of 25 people (e.g., the White Sox roster of active players) two have the same birthday. Should take the bet?

Birthday problem

Someone wants to be you $20 that in a group of 25 people (e.g., the White Sox roster of active players) two have the same birthday. Should take the bet?

Proof by Example. Pitcher Mike MacDougal (47) and infielder Paul Konerko (14) were both born on March 5.

Birthday problem

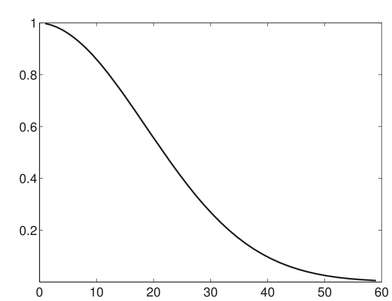
Someone wants to be you $20 that in a group of 25 people (e.g., the White Sox roster of active players) two have the same birthday. Should take the bet?

Proof by Example. Pitcher Mike MacDougal (47) and infielder Paul Konerko (14) were both born on March 5.

A match is likely because there are $(25 \cdot 24)/2 = 300$ pairs of players that share a birthday with probability $1/365$.

Probability all birthdays different for n people

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



Birthday Triples

How large a group do we need, so that three people have the same birthday?

Birthday Triples

How large a group do we need, so that three people have the same birthday?

Suppose there are 90 people and 360 birthdays. The probability no one is born on August 17 is

$$\left(1 - \frac{1}{360}\right)^{90} \approx e^{-1/4}$$

The probability k people born on that day has approximately the Poisson distribution

$$e^{-1/4}\frac{(1/4)^k}{k!} \approx 1/500 \quad \text{when } k = 3$$

P(no triple) $\approx e^{-360/500} = 0.486$. $p \approx 0.5$ for $n = 87$.

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

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

$\rightarrow$ 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

$\rightarrow$ 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.

Probability is no 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.

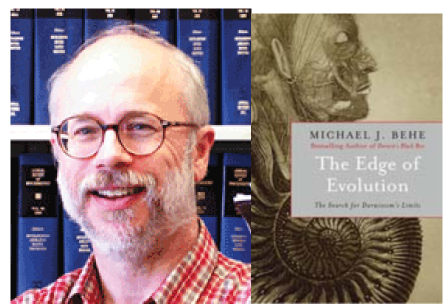
Scratch-off Triple Winner.

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

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

Michael Behe: Limits to Evolution



The Department of Biological Sciences at Lehigh University has published an official position statement which says "It is our collective position that intelligent design has no basis in science, has not been tested experimentally, and should not be regarded as scientific."

The Limits of Darwinism

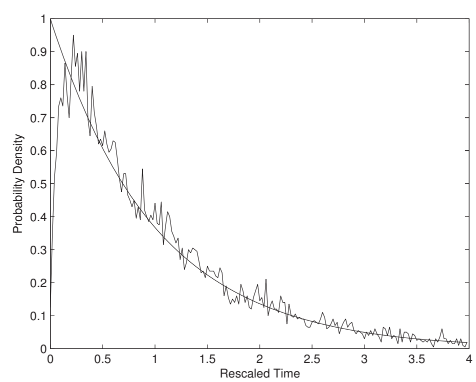
The malaria parasite *Plasmodium falciparum* has evolved resistance to chloroquine. This is due to two amino acid altering substituions in PfCRT. Michael Behe in his book *The Edge of Evolution* calls such an event a *chloroquine complexity cluster*, or CCC. He concludes:

"There are 5000 species of modern mammals. If each species had an average of a million members and if a new generation appeared every year, and if this went on for two hundred million years, the likelihood of a single CCC appearing in the whole bunch over that entire time would only be 1 in a hundred."

Theorem. If $Nu_1 \to 0$ and $N\sqrt{u_2} \to \infty$

$$P(\tau_2 > t/Nu_1\sqrt{u_2}) \to e^{-t}$$

10,000 simulations of $n = 10^3$, $u_1 = 10^{-4}$, $\sqrt{u_2} = 10^{-2}$



Behe is wrong

If $N = 10^6$, $u_1 = u_2 = 10^{-9}$, waiting time is exponential $10^{7.5} = 31.6$ million years for one prespecified pair of mutations in one species.

With 5000 species we expect this to happen in some species every 6,300 years

$u_2 \to \sqrt{u_2}$ is a factor of 31,600

Behe letter in February 2009 Genetics criticizing Durrett and Schmidt (2008) 180, 1501-1509, and our reply.

Every day 30 events of probability 1/10,000,000 happen to someone in the U.S. Of course it would be surprising if one of these happened to you.